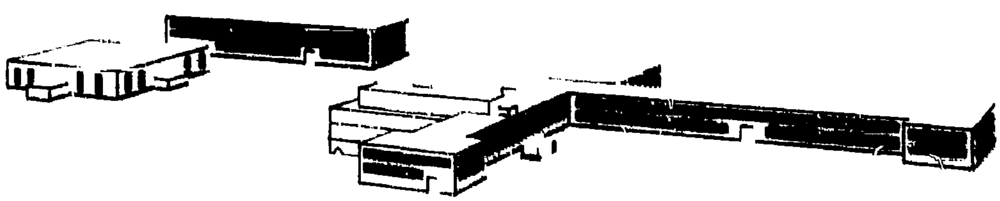
352-R-17

THE SIMULATION AND EVALUATION OF INFORMATION RETRIEVAL SYSTEMS



HRB-SINGER, INC.

SCIENCE PARK · STATE COLLEGE, PENNSYLVANIA

HRB

HRB-SINGER, INC.
A SUBSIDIARY OF THE SINGER COMPANY
Science Park, State College, Pa

352-R-17

THE SIMULATION AND EVALUATION OF INFORMATION RETRIEVAL SYSTEMS

Contract Nonr 3818(00)

April 1965

Copy No. 21 of 77 Copies

Prepared by:
C. Richard Conger

Contributor:
H. Dean Wilson

Submitted by:
K. F. Barber
Project Director

Reviewed by:
Ralph Stevenson, Manager
Methodology Branch

Approved by:
Robert Stover, Head
Behavioristics Laboratory

ABSTRACT

The assumptions, logic, and operations of an Information Retrieval System Simulation Model are described in this report. The model has been programmed in FORTRAN II for the IBM 1620. Two model variations, input and output examples, and some possibilities for improvement and development are given. Data reduction and analysis programs, operating upon the output of the simulation model, are explained and a set of sample outputs for these programs are given. Also included are listings of the latest programs developed.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

HRB

LIST OF TABLES

I. INTRODUCTION

One object of Contract Nonr 3818 (00) is the evaluation of information retrieval systems. To facilitate such evaluation, efforts are being made to develop an ideal model of these systems. The ideal model will be a simulation that will yield quantitative measurements of system performance. It will be adaptable to any specific system. For the present, the primary measurement has been the time the system requires to respond to a given request for information. Response will be either a whole or partial answer or indication that no answer is available. In addition to evaluating existing systems, the model should be useful in determining specific requirements for new information retrieval systems concepts.

The model utilization process is shown in Figure 1. Data or specifications from the system being simulated are required as inputs to the computer, as shown by example in Appendix A. The model simulation program utilizes the system input data to produce the data simulation cards, from which the simulation output is printed (Appendix B), and these cards are used as input to the Data Reduction Program and the Analysis Program (Chapter III). The output of the Data Reduction Program consists of three pages of summary statistics printed subject to several data format requirements, is illustrated in section 2, Chapter III. The Analysis Program, using simulation data cards and system acceptance criteria as inputs (Chapter III, Section 3), produces a response time estimate, a confidence level and a system acceptability decision.

At present the simulation model represents a research tool. When completed it will be useful in studying the response time aspects of computer-based information retrieval systems. Hence, in its current form its applicability to practical problems is limited. The model logic, however, can be used either to develop a simulation for a specific system or it can be extended to form a general information retrieval system simulation.

Future applications of the final simulation model should be considered now so that the subsequent model development work can be evaluated in the proper light. The model will be of greatest practical value when used as a system evaluation tool. As such, a systems engineer (user) would use it to estimate the response time of a given system. Once the user finds that the system

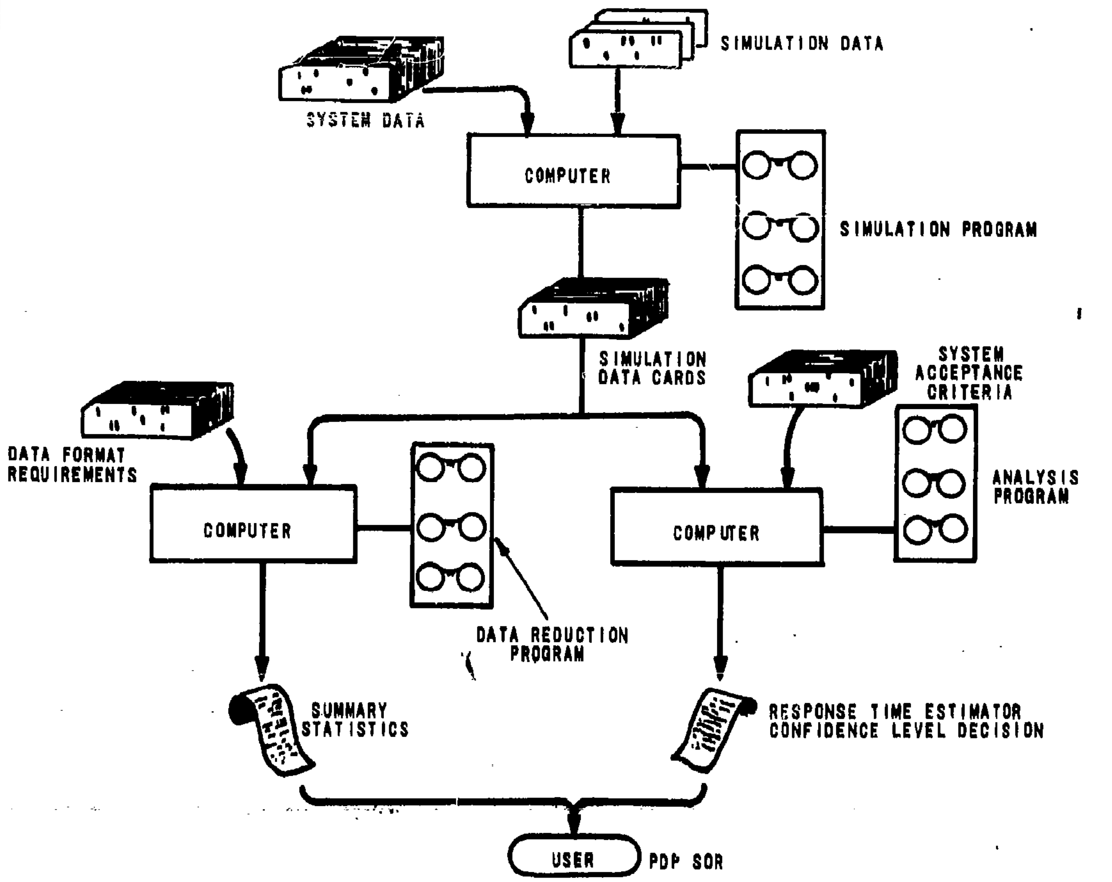


FIG. 1 - MODEL UTILIZATION PROCESS

configuration does not satisfy the predetermined set of boundary conditions, he could modify the system by changing characteristics such as the card read rate. If the system is judged acceptable after this modification, then either a new piece of equipment (e.g., with the new card read rate) must be obtained (that is, an equipment requirement has been established) or a time-equivalent improvement must be found, again by using the simulation. Once the input data for the simulation has been prepared, the changing of any particular equipment characteristics can be accomplished by simply changing the input data associated with that factor.

The simulation model requires two types of system data; namely, event-time data and selection data. In an existing system this data can be readily obtained by observation. Selection probabilities can be estimated by considering the relative frequency of use.

On the other hand, if the system to be simulated is in the concept stage of development, then it will be necessary to estimate nearly all of the input data. The system evaluator could use the model to compare two or more systems which are only "paper" systems. For example, a requirement might exist such that the response time in an acceptable system must be less than T seconds P percent of the time. To see whether any of a given set of systems satisfy this requirement, the user would simulate each (and if all were concepts he would have to estimate their corresponding parameters) and run the results either through the data reduction program (and determine the acceptance "manually") or through the analysis program. This procedure would give him a set of systems with acceptable response times. If there were a cost constraint or a space constraint (or some other limiting factor) then another selection procedure, based on these constraints, would be required.

As a research tool, the simulation model logic provides a basis for the study of operations and equipment forming an information retrieval system. The model can be extended to include a study of queueing of questions at the operator's station (and consider single channel and multiple channel service), or of real time retrieval systems where queries are queued based on a time to process (or some other) priority. The model can also be extended to include consideration of various record structures in different storage media. The motivation for studies of this sort stems from the fact that much work has been done in the area of, say, reconnaissance systems, or the obtaining of information, but we are just beginning to examine the how's and why's and where's in information retrieval systems. Many studies such as this are needed before we can be as confident in our understanding of information retrieval systems as we are of reconnaissance systems.

II. SIMULATION MODEL

One way of simulating an information retrieval system is to consider those operations which must be performed by the parts of the system. In a library, for example, certain steps are invariably followed in finding information. These steps can be called "events." Figure 2 shows the primary event sequence in a library or any other manual retrieval system. In this sequence the analyst might be a person requesting a book from a library. He would first decide how he is going to ask for which book. He would then make his request, either orally or in writing. The operator, who might be the librarian, would then interpret the question or request. The librarian might know that either the requested book was out on loan or that the library never had the book. If the operator is not sure (as happens in most cases), he would look for a loan record (after deciding where to look and what to look for).

This event sequence would continue until the analyst (or book requester) receives an answer which might be either the requested book, a different book (a mistake), or a statement that the book is not currently available. Both the first and third answers would be acceptable. The second answer would not be acceptable, and the process might be repeated again and again until an acceptable answer is received.

When the number of requests entered into the system becomes large or the files to be searched become unmanageable, the manual retrieval system would ordinarily be supplanted by a computer-based information retrieval system. Several new events would now appear in the primary event sequence, as seen in Figure 3. The operator in Figure 3 is acting as an interpreter (between analyst and computer). No longer can he be simply the traditional librarian. Now he must consider how he is to ask the computer to do his searching, how he is to tell the computer of his request, how he wants his answer, and how he is to receive the answer.

The operator may also have a choice to make in asking for information. In this report a distinction is made between a question and a query. A question is a request from the analyst while a query is a request by the operator. The operator may need to use various types of queries at various times, such as

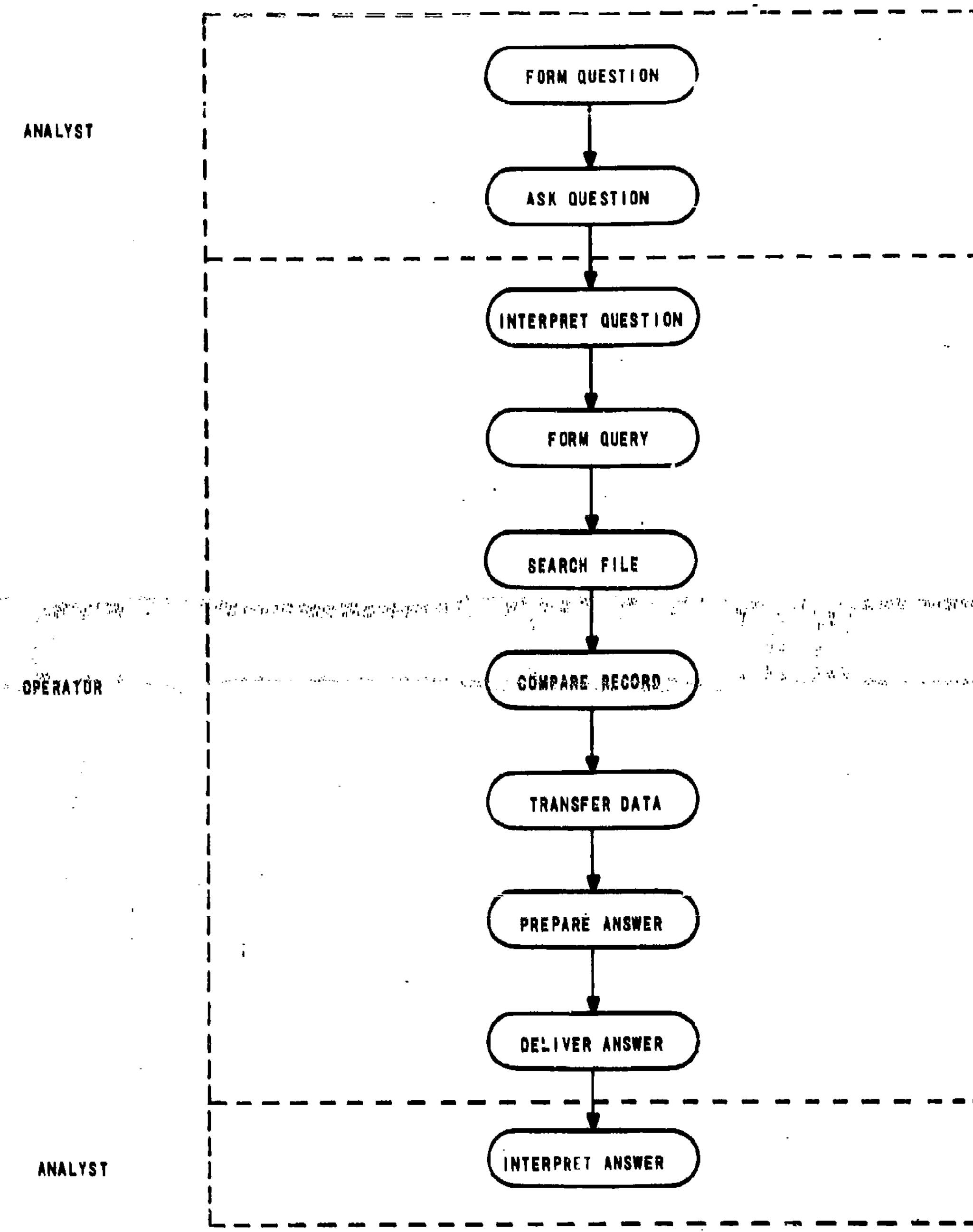


FIG. 2 - PRIMARY EVENT SEQUENCE - MANUAL RETRIEVAL SYSTEM

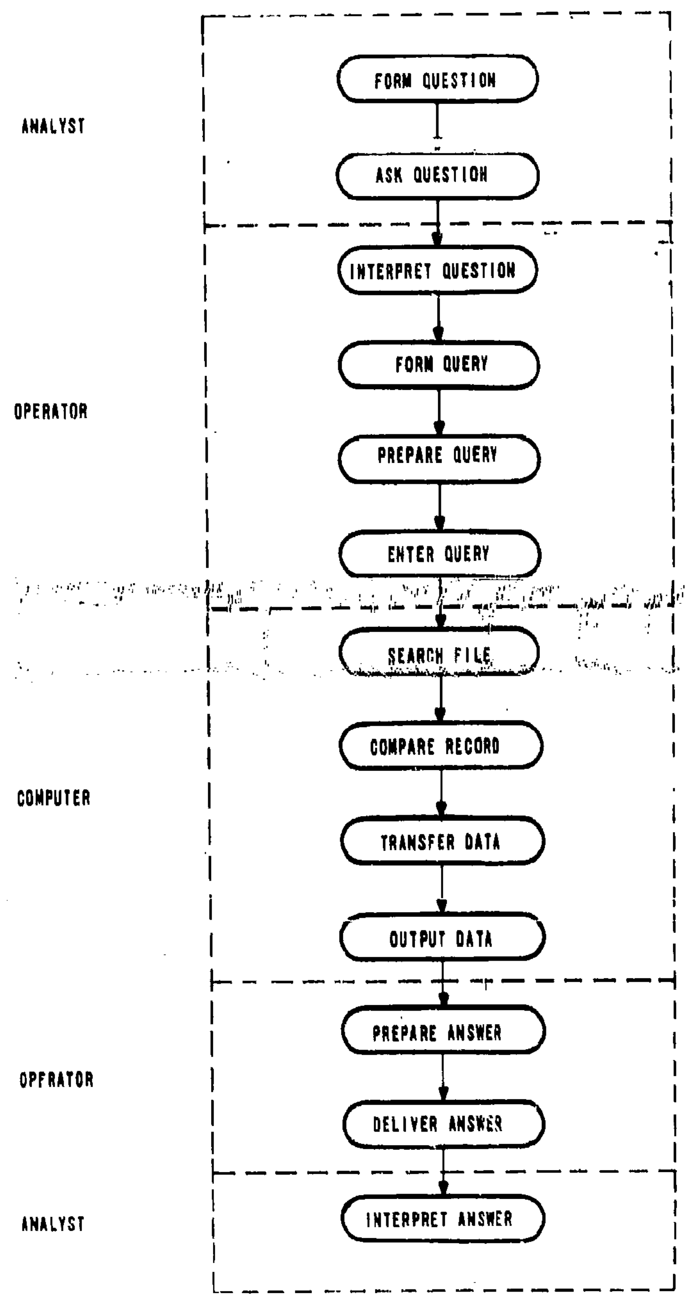


FIG 3 - PRIMARY EVENT SEQUENCE - COMPUTER RETRIEVAL SYSTEM

comparison, summaries and extremum. For a given question, he may require several queries of each type in order to satisfy the request. Suppose, for example, that the information in a record has been stored in four groups:

1.	Person's name	(field 1)
2.	Service number	(field 2)
3.	Rank	(field 3)
4.	Station	(field 4)

An analyst might want to know, for example, to what ship an individual named CONGER is assigned. Following the procedure suggested in Figure 3, the analyst would ask the operator to obtain the desired information from the system. The operator in turn might fill out a number of query forms to elicit the following information:

1. CONGER
2. 7500743
3. RM2
4. USS ANTIETAM

In a query, similar to the one below, an X indicates the information desired and a blank means the information is not wanted.

1.	CONGER	
2.	-	
3.	-	
4.	X	
5.	1	(query type: comparison)

On the other hand the analyst may want to know the names of all the personnel aboard the USS ANTIETAM, CVS-36. In this case the query form might look like

1.	X	
2.	-	
3.	-	
4.	USS ANTIETAM	
5.	2	(query type: summary)

More than one query can be generated by a question since it is possible to consider the analyst asking, for example, whether CONGER's serial number is 7500742. In this case the operator may form two queries, the first being

1. CONGER
2. X
3. -
4. -
5. 1

and the second

1. X
2. 7500742
3. -
4. -
5. 1

Questions such as "What ship has the most personnel aboard?" would cause an extremum query type to be generated. With such a query, all ships would be examined, the number of personnel aboard each ship determined, and the ship with the largest number printed out as the answer.

The examples used are given purely for illustration of query types. The model will provide for use of up to 10 query types by the system being simulated. It is also assumed that each question asked by the analyst can be handled differently by the system; that is, that it would be possible to establish "question categories" and within each find that the system responds in essentially the same way.

The basic model logic is centered on the response time measure. In essence, the response time is equal to the sum of all of the event times and associated delays. The events in Figure 4 can be defined as follows:

FORM QUESTION	Determining what must be asked and how it is to be asked
ASK QUESTION	The process of asking the question
INTERPRET QUESTION	Determining whether this question can be answered at this time

DETERMINE NUMBER OF QUERIES	Determining how many queries will be necessary to answer this question
FORM QUERY	Determining what the query must say
PREPARE QUERY	The process of preparing the query
ENTER QUERY	The process of entering the query
SEARCH FILE	Examining the data base
TRANSFER RECORD	Shifting data prior to output
OUTPUT DATA	Outputting selected data
PREPARE ANSWER	Collecting output data relevant to the given question
DELIVER ANSWER	The process of delivery of the answer
INTERCEPT ANSWER	Comparing the given answer with the expected answer
INFORM ANALYST	Refusing a question
REPHRASE QUESTION	Determining how to ask what must be asked

With the exception of the fourth event, all of these are basic time events; that is, the time it takes to perform the particular event will contribute to the overall time it takes to respond to the given request (response time).

Some of the contingency cases have been included in Figure 4. The first contingency occurs when the request (question) posed by the analyst is not acceptable. In this case the operator so informs the analyst, who then reforms his question and again makes a request for information.

The second contingency occurs when the query is rejected by the system logic; for example, when a simple coding error has been detected. This part of the basic model logic can be expanded to include both query checking and information checking (determining whether the system has any data on the requested subject).

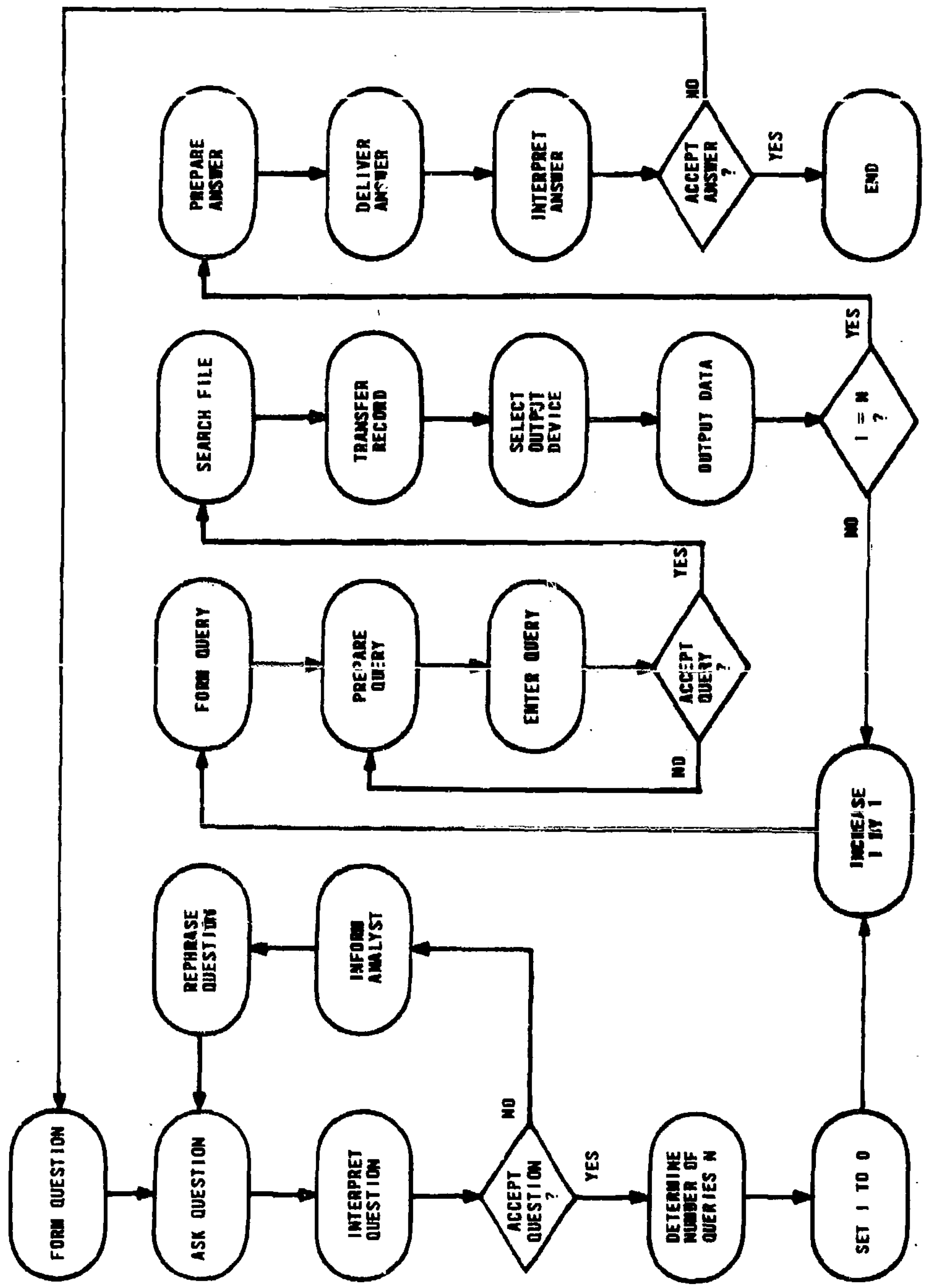


FIG. 4 - BASIC MODEL LOGIC

The third contingency occurs when the answer is rejected by the analyst. Even though an answer has been received, as shown in Figure 4, the simulation does not stop until the answer is acceptable, which means that the time addition continues until END is reached.

1. MOD II

Based on the logic shown in Figure 4, two models have been developed, programmed, andrun. These models differ in the way the number of queries is selected as well as in the way the query types are selected. In the first model (Mod I) a random number of queries of each type aro selected; in the second model (Mod II) the number of queries to be considered is selected first, and then for each query a query type is selected (again at random according to some selection rule).

Figure 5 shows the logic followed in Mod II. It has been assumed that a particular time event will take anywhere from some time T_1 to some time T_2; that is, it will take no less than T_1 and no more than T_2. Each event can have its own time range. In the simulation all times are measured in seconds. Rather than bogging down in a detailed problem in the early stage of model development, we have assumed that the probability of each time event can be represented by a uniform distribution, which is analogous (in the discrete case) to saying that if a ~~die is~~ rolled and a 1 comes up, the event took a minimum time, ~~or if a 2~~ comes up, the event took T seconds longer, or, finally, if a 6 comes up, the event took a maximum time (also assuming an unloaded die). If a uniform time distribution is used and if T_1 represents the minimum time and T_2 represents the maximum time, then the procedure for selection of a random event time within this interval is to pick a random number R in the interval from 0 to 1 and then substitute it in the following equation, where T represents the selected event time:

$$T = T_1 + (T_2 - T_1) \cdot R. \qquad (1)$$

Figure 6 shows the Mod II flow chart. Whenever "SELECT T_i" is required, equation 1 i applied by picking R at random and then substituting corresponding min and max event times in the equation. In this flow chart will also been seen a statement

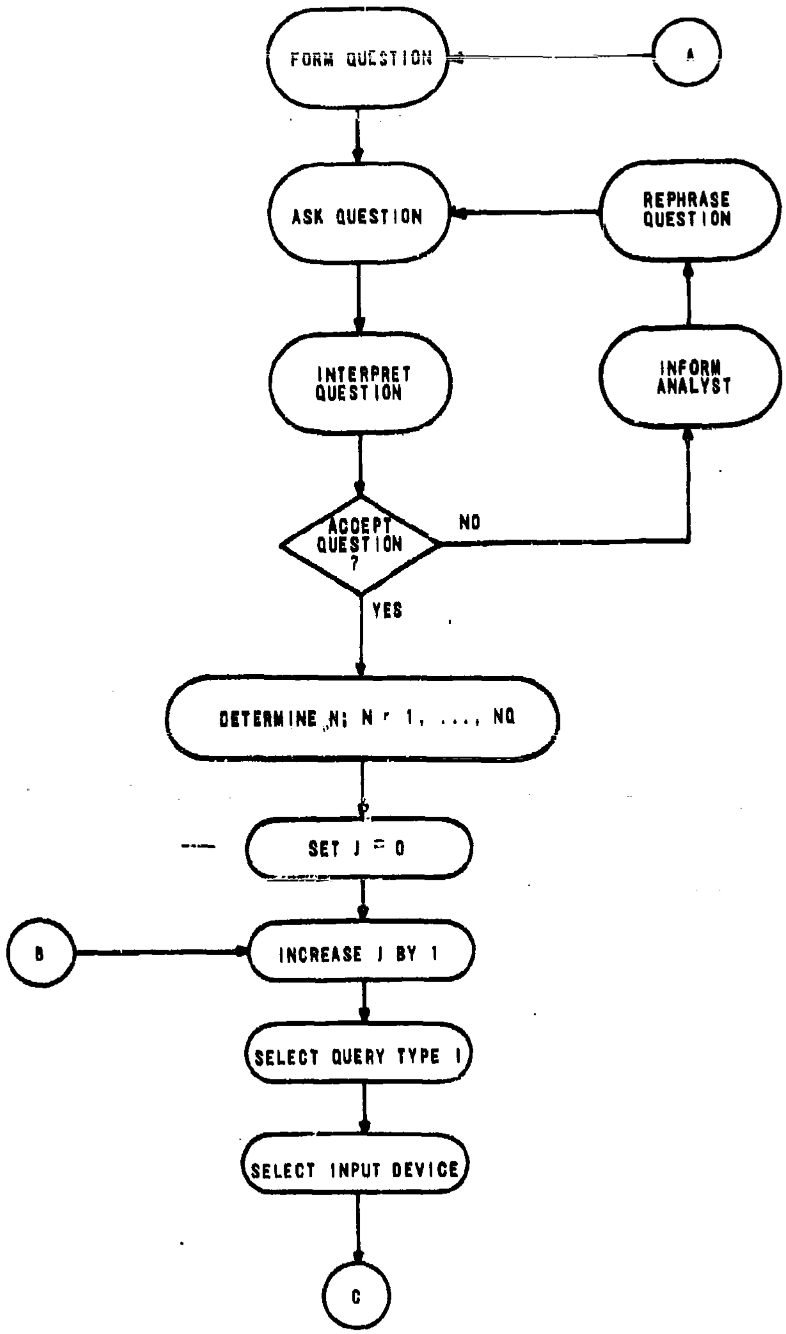


FIG. 5 - MOD II LOGIC

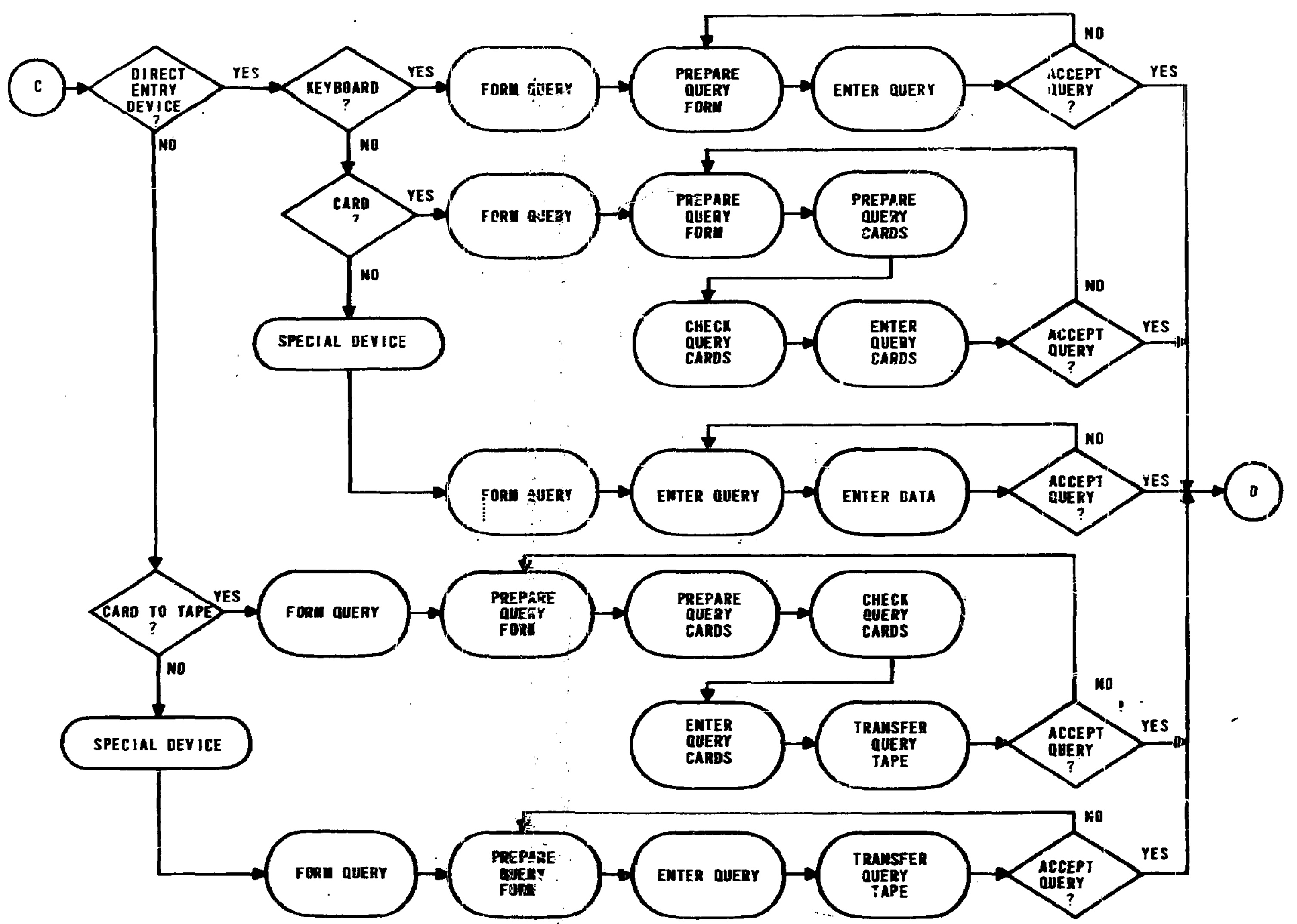


FIG. 5 - MOD II LOGIC (CONT'D)

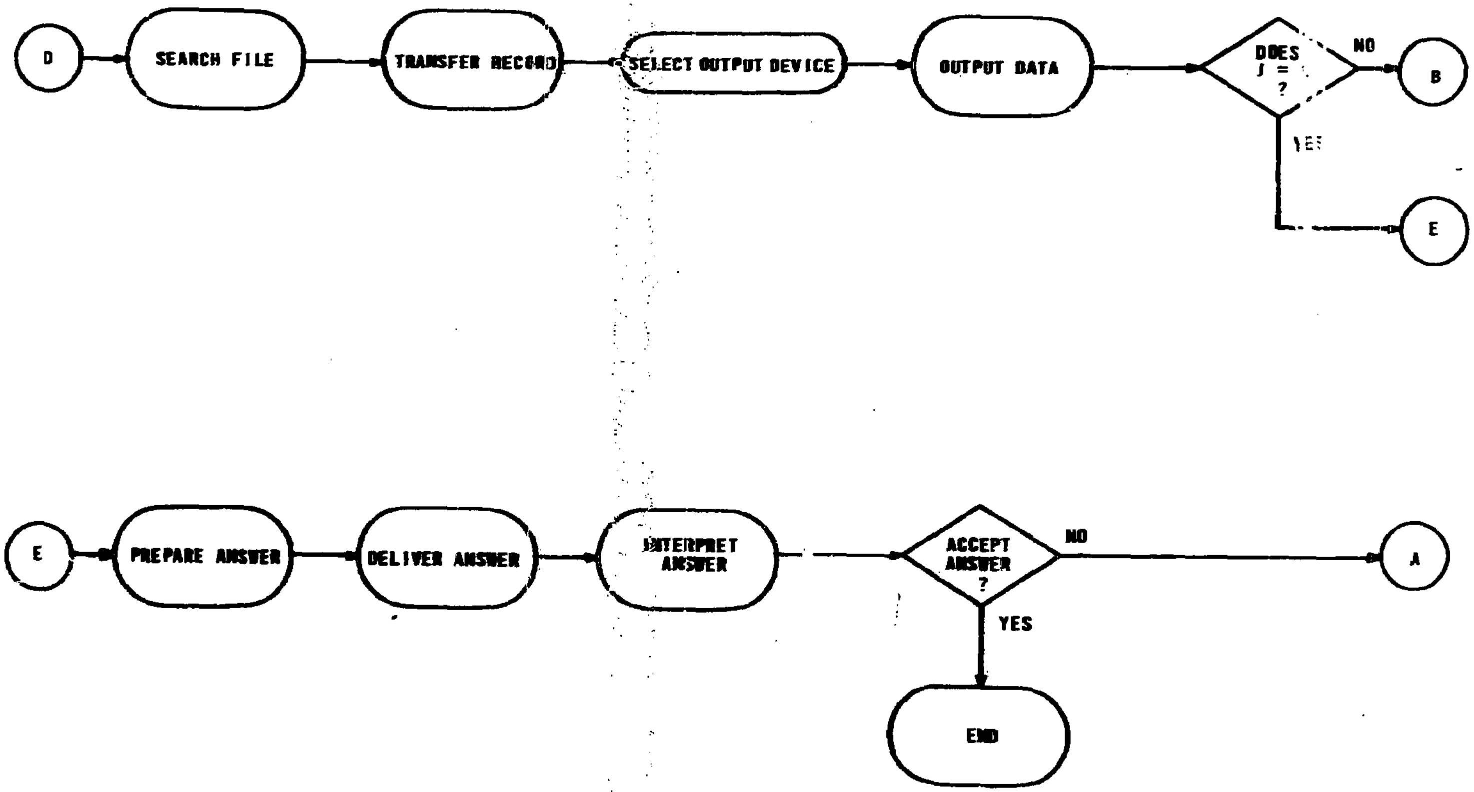


FIG. 5 - MOD II LOGIC (CONT'D)

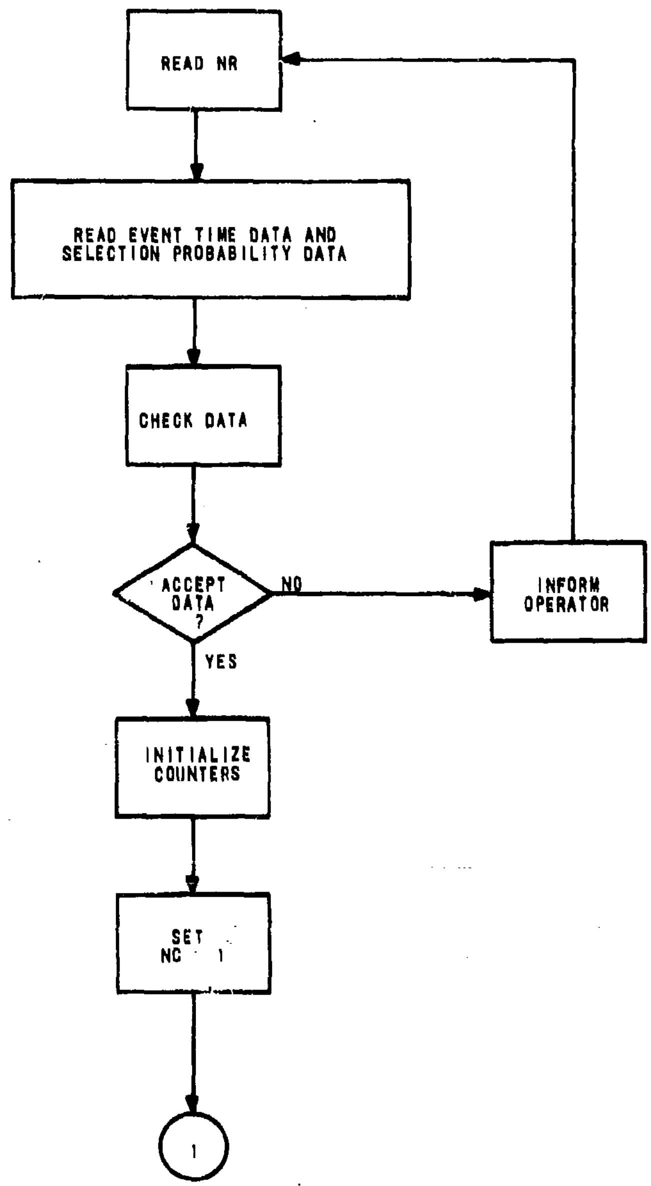


FIG. 6 - MOD II FLOW CHART

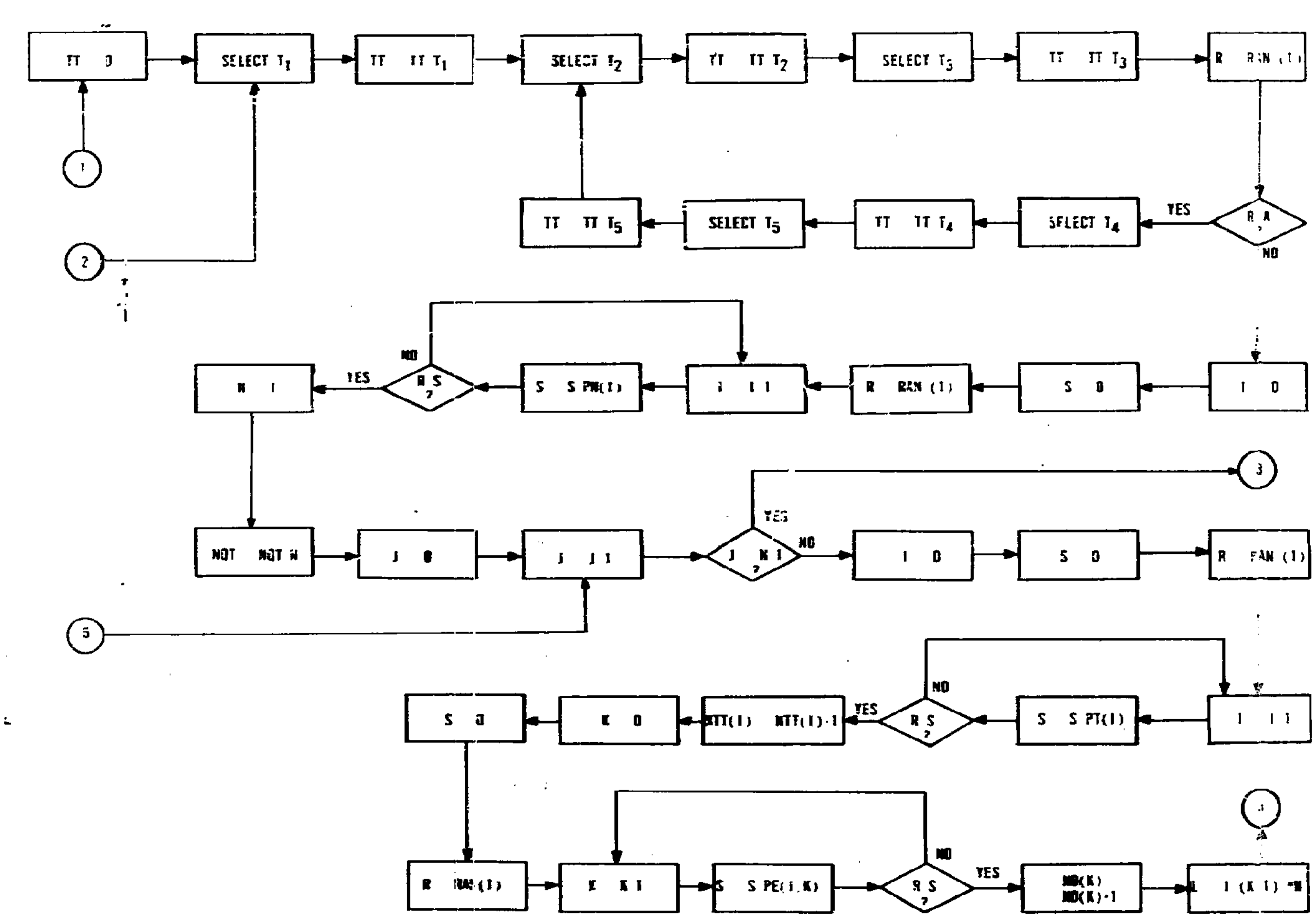


FIG. 6 - MOD II FLOW CHART (CONT'D)

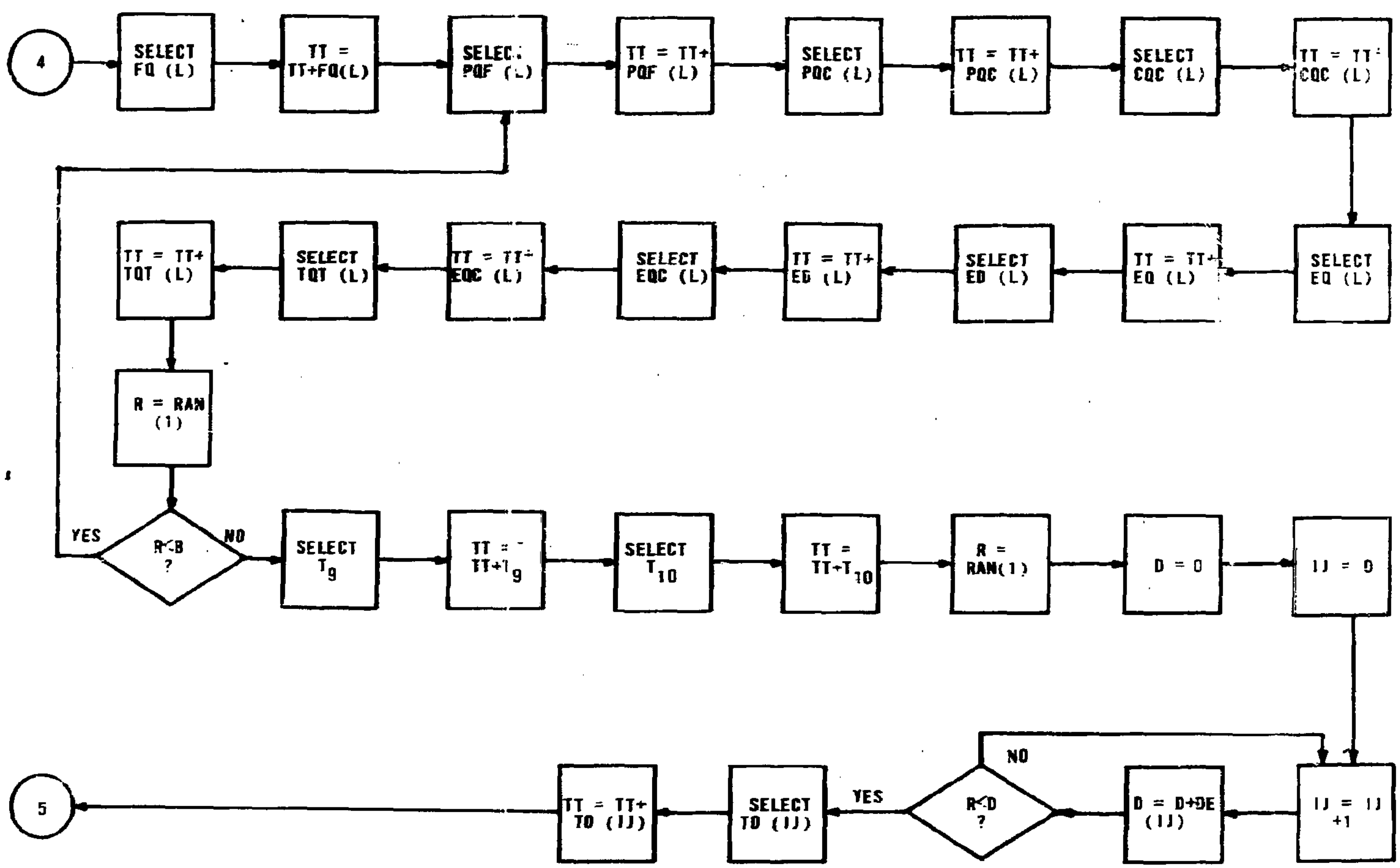


FIG. 6 - BOB II FLOW CHART (CONT'D)

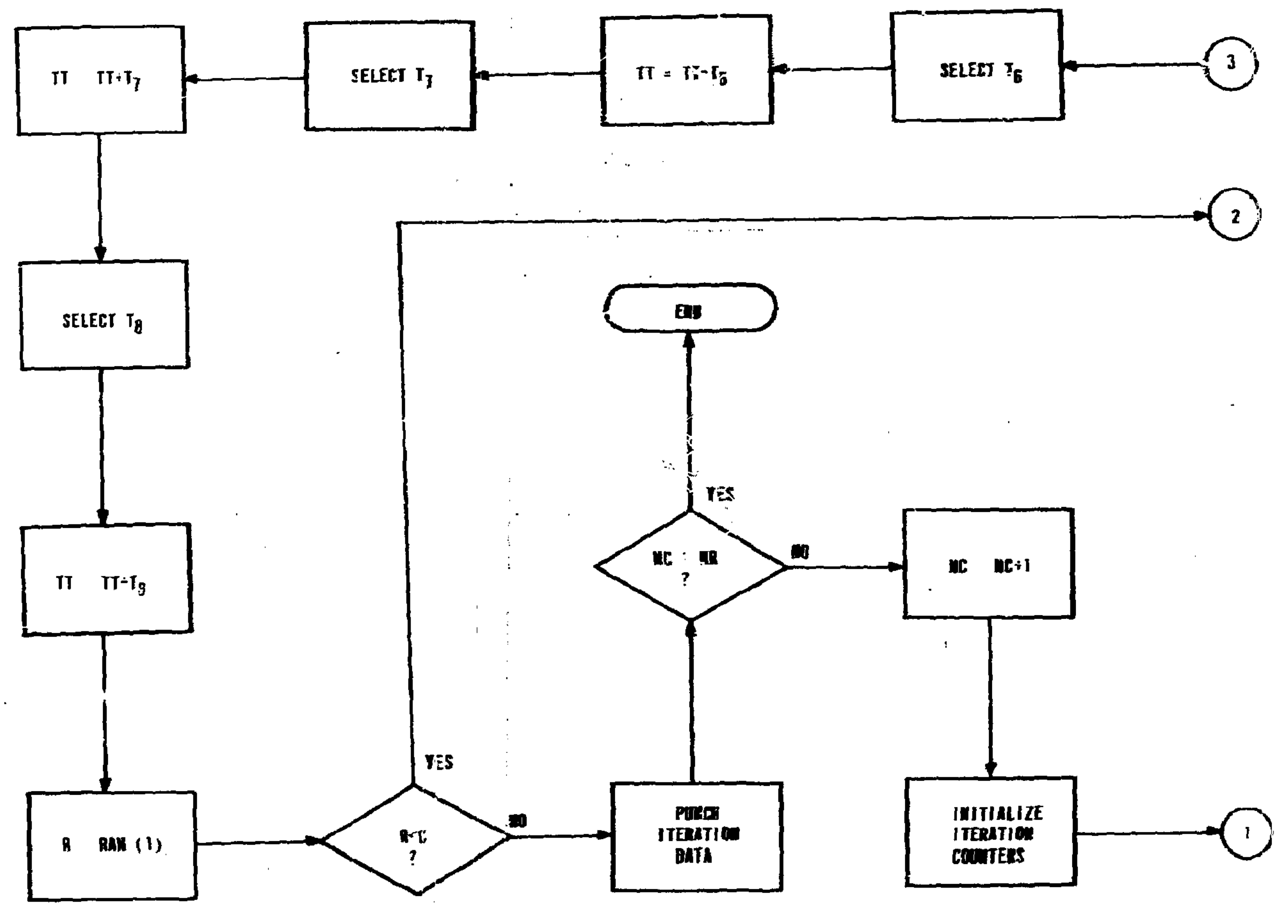


FIG. 6 - MOD II FLOW CHART (CONT'D)

$$R = RAN(1.)$$

This statement calls for a subroutine which will generate a uniformly distributed random number and then set the value of R equal to this number. A different random number will be generated every time this subroutine is used.

Table 1 shows that the event sequence in the model differs according to the type of entry device selected. For example, the system being simulated may have a special query entry device that enables the operator to enter a query simply by pushing a button on a console (this would indicate to the computer that a certain query type is desired) and then entering the specific data. In this case the operator first forms his query (FORM QUERY), then pushes a button (ENTER QUERY), and then enters his data (ENTER DATA). Specific query types may be entered by certain entry devices more often than by others (assuming a choice). The simulation allows for this preference as shown in Table 2.

TABLE 1 - EVENT SEQUENCE VERSUS INPUT DEVICE

	DIRECT ENTRY			INDIRECT ENTRY	
	KEYBOARD	CARD	SPECIAL DEVICE	CARD AND TAPE	SPECIAL DEVICE
FORM QUERY	X	X	X	X	X
PREPARE QUERY FORM	X	X		X	X
PREPARE QUERY CARDS		X		X	
CHECK QUERY CARDS		X		X	
ENTER QUERY	X		X		X
ENTER DATA			X		
ENTER QUERY CARDS		X		X	
TRANSFER QUERY TAPE				X	X

NOTE: AN X INDICATES THAT THE PARTICULAR EVENT WOULD OCCUR IF THAT INPUT DEVICE WERE USED.

TABLE 2. INPUT DEVICE SELECTION PROBABILITY MATRIX

Query Type	Input Device Number 1	2	. . .	N
Q_1	P_{11}	P_{12}	. . .	P_{1N}
Q_2	P_{21}	P_{22}	. . .	P_{2N}
$\vdots$	$\vdots$	$\vdots$	. . .	$\vdots$
Q_M	P_{M1}	P_{M2}	. . .	P_{MN}

NOTE: (1) P_{ij} is the conditional probability of selecting input device j, given that query type i has been selected;

(2) N is the number of input devices in the system;

(3) M is the number of query categories (types).

It was mentioned earlier that Mod II differs just slightly from Mod I. The major difference is in the selection of the number of queries and the selection of query types. In Mod II, the number of queries to be considered in a particular run is selected first. Then for the first query, a query type is picked according to a specified selection procedure, and this query is "run." If there are more queries to be generated, then for the second query a query type is picked. In this "picking" sampling with replacement is performed, thereby giving a constant selection probability for each query type.

There is a slight difference in the Mod II output. Since Mod I was primarily a prototype, the program for Mod II includes several refinements not found in Mod I. Mod II will now act as a prototype for a subsequent model. It is for this reason that only Mod II is described in detail in this report.

The simulation requires two types of inputs: event times and selection probabilities. Event time data describes the time range for a given event. For example, it might take the operator anywhere from 5 to 15 seconds to interpret a question given to him by an analyst. Selection probability data refers to the

observed usage of the various query types and I/O devices. For example, it might have been observed that 60% of the time an on-line printer was used for output of data, 20% of the time a console typewriter was used, and 20% of the time an off-line printer was used.

Specific input requirements can be seen in the program listing while a numerical input example is given in Appendix A.

TABLE 3. MOD II SYMBOL TABLE

Symbol	Meaning
A	Probability of rejecting a question
B	Probability of rejecting a query
C	Probability of rejecting an answer
D	Dummy variable
I	Dummy variable
IJ	Dummy variable
J	Dummy variable
K	Dummy variable
L	Dummy variable
M	Number of query types
NC	Iteration number
NI	Number of input devices
NO	Number of output devices
NQ	Number of queries (maximum)
NR	Number of iterations for the run
R	Random number dummy variable
RAN(1.)	Uniform random number generator call expression
S	Dummy variable
TT	Total time

TABLE 3. MOD II SYMBOL TABLE (Cont'd)	
Symbol	Meaning
CQC (L, 1)	Minimum time to check I^{th} query cards (CQC) for K^{th} device, where L = I + (K-1) M
CQC (L, 2)	Maximum time to check I^{th} query cards for K^{th} device
DE (I)	Probability of selecting I^{th} output device
ED (L, 1)	Minimum time to enter I^{th} query data (ED) for K^{th} device
ED (L, 2)	Maximum time to enter I^{th} query data (ED) for K^{th} device
EQ (L, 1)	Minimum time to enter I^{th} query (EQ) for K^{th} device
EQ (L, 2)	Maximum time to enter I^{th} query (EQ) for K^{th} device
EQC (L, 1)	Minimum time to enter I^{th} query cards (EQC) for K^{th} device
EQC (L, 2)	Maximum time to enter I^{th} query cards (EQC) for K^{th} device
FQ (L, 1)	Minimum time to form I^{th} query (FQ) for K^{th} device
FQ (L, 2)	Maximum time to form I^{th} query (FQ) for K^{th} device
N (I)	Number of queries of I^{th} type
PE (I, J)	Probability of selecting J^{th} input device for I^{th} query type
PN (I)	Probability of I queries being selected, where I = 1, 2, . . . , NQ
PQC (L, 1)	Minimum time to prepare I^{th} query cards (PQC) for K^{th} device
PQC (L, 2)	Maximum time to prepare I^{th} query cards (PQC) for K^{th} device
PQF (L, 1)	Minimum time to prepare I^{th} query form (PQF) K^{th} device

TABLE 3. MOD II SYMBOL TABLE (Cont'd)	
Symbol	**Meaning**
PQF (L, 2)	Maximum time to prepare I^{th} query form (PQF) for K^{th} device
PT (I)	Probability of selecting query type I
T (I, 1)	Minimum time for I^{th} event
T (I, 2)	Maximum time for I^{th} event
TO (I, 1)	Minimum output time on I^{th} device
TO (I, 2)	Maximum output time on I^{th} device
TQT (L, 1)	Minimum time to transfer I^{th} query tape (TQT) for K^{th} device
TQT (L, 2)	Maximum time to transfer I^{th} query tape (TQT) for K^{th} device

2. EXAMPLE

To show how the simulation model works, assume, for example, a system which has the following characteristics:

a. fixed user group;
b. fixed data type (one kind of data in a record);
c. five query catagories;
d. magnetic tape data storage;
e. fixed record structure;
f. input devices: card to core and keyboard to core;
g. output devices: on-line printer, off-line printer, and console typewriter.

Assume also that the method of operation is essentially as illustrated in Figure 4; that is, a question form is prepared by the user and given to the operator. The operator then interprets the question, and if the question is acceptable he prepares a query form. The query is then entered into the system. The output of the run is given to the user. The contingency cases shown in Figure 4 are also assumed to hold.

Table 4 (Appendix A) shows the input data, which is just a numerical sample -- not actual system data, necessary for Mod II.

```
C     INFORMATION RETRIEVAL SYSTEM SIMULATION
C     MOD 2
C     PROGRAM LIMITS-10 QUERY TYPES, 5 INPUT DEVICES, 5 OUTPUT DEVICES
C     10 QUERIES PER QUESTION BASIC MAXIMUM
C
      DIMENSION T(10,2), PN(10), PT(10), PE(10,5), FQ(50,2),
     1PQF(50,2),PQC(50,2),CQC(50,2),EQ(50,2),ED(50,2),EQC(50,2),
     2TQT(50,2),DE(5),NTT(10),TO(5,2),ND(10)
  100 FORMAT(I5)
  101 FORMAT(4I3)
  102 FORMAT(2F10.4)
  103 FORMAT(10F6.3)
  104 FORMAT(10F6.3)
  105 FORMAT(5F6.3)
  106 FORMAT(2F10.4)
  107 FORMAT(2F10.4)
  108 FORMAT(2F10.4)
  109 FORMAT(2F10.4)
  110 FORMAT(2F10.4)
  111 FORMAT(2F10.4)
  112 FORMAT(2F10.4)
  113 FORMAT(2F10.4)
  114 FORMAT(F6.3,2F10.4)
  115 FORMAT(3F6.3)
  200 FORMAT(I4,F10.2,I4,10I3,10I3)
  300 FORMAT(10HDATA ERROR)
  301 FORMAT(42HCORRECT AND RE-ENTER ALL DATA - PUSH START)
    1 READ 100,NR
      READ 101, M, NI, NO, NQ
      DO 2 I=1,10
      READ 102, T(I,1), T(I,2)
      IF(T(I,1)-T(I,2))2,2,900
    2 CONTINUE
      S=0
      READ 103, (PN(I),I=1,NQ)
      DO 3 I=1,NQ
    3 S=S+PN(I)
      IF(S-1.)900,4,900
    4 S=0
      READ 104,(PT(I),I=1,M)
      DO 5 I=1,M
    5 S=S+PT(I)
      IF(S-1.)900,6,900
    6 DO 8 I=1,M
      READ 105,(PE(I,J),J=1,NI)
      S=0
      DO 7 J=1,NI
    7 S=S+PE(I,J)
      IF(S-1.)900,8,900
    8 CONTINUE
```

```
      L=M*NI
      DO 9 I=1,L
      READ 106,FQ(I,1),FQ(I,2)
      IF(FQ(I,1)-FQ(I,2))9,9,900
    9 CONTINUE
      DO 10 I=1,L
      READ 107,PQF(I,1),PQF(I,2)
      IF(PQF(I,1)-PQF(I,2))10,10,900
   10 CONTINUE
      DO 11 I=1,L
      READ 108,PQC(I,1),PQC(I,2)
      IF(PQC(I,1)-PQC(I,2))11,11,900
   11 CONTINUE
      DO 12 I=1,L
      READ 109,CQC(I,1),CQC(I,2)
      IF(CQC(I,1)-CQC(I,2))12,12,900
   12 CONTINUE
      DO 13 I=1,L
      READ 110,EQ(I,1),EQ(I,2)
      IF(EQ(I,1)-EQ(I,2))13,13,900
   13 CONTINUE
      DO 14 I=1,L
      READ 111,ED(I,1),ED(I,2)
      IF(ED(I,1)-ED(I,2))14,14,900
   14 CONTINUE
      DO 15 I=1,L
      READ 112,EQC(I,1),EQC(I,2)
      IF(EQC(I,1)-EQC(I,2))15,15,900
   15 CONTINUE
      DO 16 I=1,L
      READ 113,TQT(I,1),TQT(I,2)
      IF(TQT(I,1)-TQT(I,2))16,16,900
   16 CONTINUE
      S=0
      DO 17 I=1,NO
      READ 114,DE(I),TO(I,1),TO(I,2)
      IF(TO(I,1)-TO(I,2))17,17,900
   17 S=S+DE(I)
      IF(S-1.)900,18,900
   18 READ 115,A,B,C
      IF(A-1.)19,900,900
   19 IF(B-1.)20,900,900
   20 IF(C-1.)21,900,900
   21 NC=0
      KK=NI+NO
   22 IT=0
      NC=NC+1
      DO 23 I=1,KK
   23 ND(I)=0
      NQT=0
```

```
      DO 24 I=1,M
   24 NTT(I)=0
   25 TT=TT+T(1,1)+(T(1,2)-T(1,1))*RAN(1.)
   26 TT=TT+T(2,1)+(T(2,2)-T(2,1))*RAN(1.)
      TT=TT+T(3,1)+(T(3,2)-T(3,1))*RAN(1.)
      R=RAN(1.)
      IF(R-A)27,28,28
   27 TT=TT+T(4,1)+(T(4,2)-T(4,1))*RAN(1.)
      TT=TT+T(5,1)+(T(5,2)-T(5,1))*RAN(1.)
      GO TO 26
   28 I=0
      S=0
      R=RAN(1.)
   29 I=I+1
      S=S+PN(I)
      IF(R-S)30,29,29
   30 N=I
      NQT=NQT+N
      J=0
   31 J=J+1
      IF(J-N-1)34,32,900
   32 TT=TT+T(6,1)+(T(6,2)-T(6,1))*RAN(1.)
      TT=TT+T(7,1)+(T(7,2)-T(7,1))*RAN(1.)
      TT=TT+T(8,1)+(T(8,2)-T(8,1))*RAN(1.)
      PUNCH 200,NC,TT,NQT,(NTT(I),I=1,M),(ND(I),I=1,KK)
      R=RAN(1.)
      IF(R-C)25,33,33
   33 IF(NC-NR)22,901,901
   34 I=0
      S=0
      R=RAN(1.)
   35 I=I+1
      S=S+PT(I)
      IF(R-S)36,35,35
   36 NTT(I)=NTT(I)+1
      K=0
      S=0
      R=RAN(1.)
   37 K=K+1
      S=S+PE(I,K)
      IF(R-S)38,37,37
   38 ND(K)=ND(K)+1
      L=I+(K-1)*M
      TT=TT+FQ(L,1)+(FQ(L,2)-FQ(L,1))*RAN(1.)
   39 TT=TT+PQF(L,1)+(PQF(L,2)-PQF(L,1))*RAN(1.)
      TT=TT+PQC(L,1)+(PQC(L,2)-PQC(L,1))*RAN(1.)
      TT=TT+CQC(L,1)+(CQC(L,2)-CQC(L,1))*RAN(1.)
      TT=TT+EQ(L,1)+(EQ(L,2)-EQ(L,1))*RAN(1.)
      TT=TT+ED(L,1)+(ED(L,2)-ED(L,1))*RAN(1.)
      TT=TT+EQC(L,1)+(EQC(L,2)-EQC(L,1))*RAN(1.)
```

```
      TT=TT+TQT(L,1)+(TQT(L,2)-TQT(L,1))*RAN(1.)
      R=RAN(1.)
      IF(R-B)39,40,40
   40 TT=TT+T(9,1)+(T(9,2)-T(9,1))*RAN(1.)
      TT=TT+T(10,1)+(T(10,2)-T(10,1))*RAN(1.)
      R=RAN(1.)
      S=0
      I=0
   41 I=I+1
      S=S+DF(I)
      IF(R-S)42,41,41
   42 TT=TT+TO(I,1)+(TO(I,2)-TO(I,1))*RAN(1.)
      I=I+NI
      ND(I)=ND(I)+1
      GO TO 31
  900 TYPE 300
      TYPE 301
  901 PAUSE
      GO TO 1
      END
```

For this example, the I/O devices are labeled as follows:

Input Device	Output Device
1 - card	1 - on-line printer
2 - keyboard	2 - off-line printer
	3 - console typewriter

Since both input devices are of the direct entry type, the ENTER DATA (ED) and the TRANSFER QUERY TAPE (TQT) events do not apply, and it will be necessary to insert 20 blank cards for the ED matrix and 20 blank cards for the TQT matrix.

3. MODEL DEVELOPMENT

Two factors will be incorporated as the model develops. The first is equipment characteristics, and the second is a requirement for more freedom in specifying time data.

Equipment Characteristics

At present Mod II does no more than approximate the actual operation of the equipment in a system. For example, in entering a query, only the time range is specified. For the model to be an effective evaluation tool, it should possess a capability for assessing variations in equipment, say, in the read rate of a card reader so that the sensitivity of the response time could be studied in light of this variation. If all other factors were fixed, then a threshold could be determined for this variable (read rate).

The incorporation of equipment characteristics in the model does present several problems. For example, consider the factors involved with the entry of data (a query) by way of a direct entry keyboard. This event can be described by at least three parameters: entry rate, data volume, and data form factor. The entry rate will be an equipment characteristic which will be the maximum number of characters that can be entered per second, and hence will be measured in characters per second. Data volume will be a function of the type of query, and it seems reasonable to assume that there will be a distribution of characters for each type of query. Data volume, then, will be measured in characters. The

data form factor will relate to the complexity of the data, and it could be assumed that this factor will remain constant for each type of query. The data form factor will be a dimensionless quantity. If R_e represents the entry rate, $V(Q)$ the data volume of query Q, and $F(Q)$ the data form factor for query Q, then the entry time, expressed in seconds, is given by

$$T_e(Q) = \frac{F(Q)\,V(Q)}{R_e}.$$

We can represent an adjusted input rate by $R_e^{\prime}(Q) = R_e/F(Q)$, which shows the dependence of the input rate on the ability of an operator to enter data variations resulting from query complexity. The best case, $R_e^{\prime} = R_e$, occurs when the degree of complexity is minimum. The maximum of $F(Q)$ could be such that $R_e^{\prime}(Q)$ equals the "hunt and peck" rate. An additional assumption can be made about $F(Q)$: that there is sufficient volume of data in each query type to make the factor meaningful.

The "enter query cards" event can be handled in a similar way. In this case only two factors are needed: entry rate and data volume. The entry rate would be the maximum number of cards which can be entered per second, and hence measured in cards per second. The data volume factor would be a function of the query type, and would be measured in cards. It can be assumed that each query type will have associated with it a discrete distribution for the number of cards necessary. If again R_e represents the entry rate and $V(Q)$ the data volume of query type Q, then the entry time, expressed in seconds, is given by

$$T_e(Q) = V(Q)/R_e.$$

It can be seen that by including equipment characteristics, more significant details of the system are included automatically, thereby making the model potentially more effective and the results more reliable.

Specifying Time Data

In the present simulation model all time factors are uniformly distributed random variables -- the only allowance being that the domain can be an arbitrary interval of time. For the actual time events, however, a uniform distribution

would really be a poor choice for an approximating probability density function (pdf). Since the time data used in the simulation will represent observations of the actual events (in most cases) and these data will be used to develop a histogram (time vs. number of occurrences -- see Figure 8) of the event, then it would seem natural to approximate this histogram to represent the event pdf. The choice of the type of approximating curve is dependent on the degree of accuracy desired. Confidence in the final results can at least be improved by using a line segment approximation to the parts of the histogram (see Figure 7, overlay). Approximation of the time distributions of the various events by sets of line segment shows that the current nodel (with its uniformly distributed random variables) is a special case.

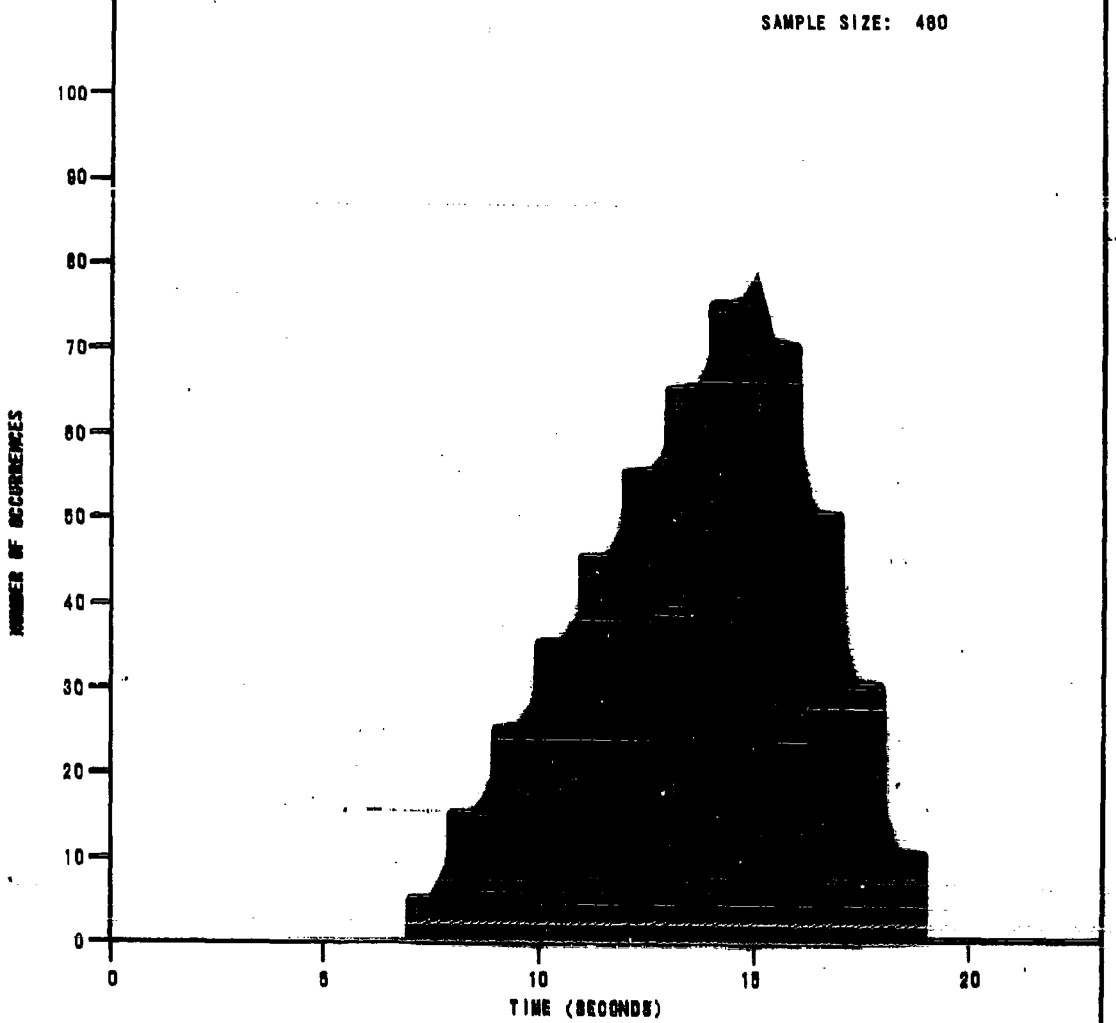
The following pages discuss the derivation of an equation which will provide a one-to-one transformation from a uniformly distributed random variable to a variable ..aving a pdf approximated by a given set of line segments. (Most computer installations, including the IBM 1620 at HRB-Singer, have only this subroutine for the generation of random numbers.) Also included is a listing of the random number generator subroutine (which will be used in a subsequent simulation model) along with a listing of a check program for this subroutine.

The following procedure can be adopted to prepare event time input data for the subsequent simulation model:

1. Samplo the event -- that is, time the event from the start to its finish and record; repeat n - 1 times for a sample of size n;
2. Construct a histogram of the number of occurrences vs. time;
3. Draw an approximating curve (composed of line segments);
4. List end points of each line segment (ordered by increasing values of of the time coordinate).

For example, suppose that a particular event is observed 480 times and it is found that the following is true:

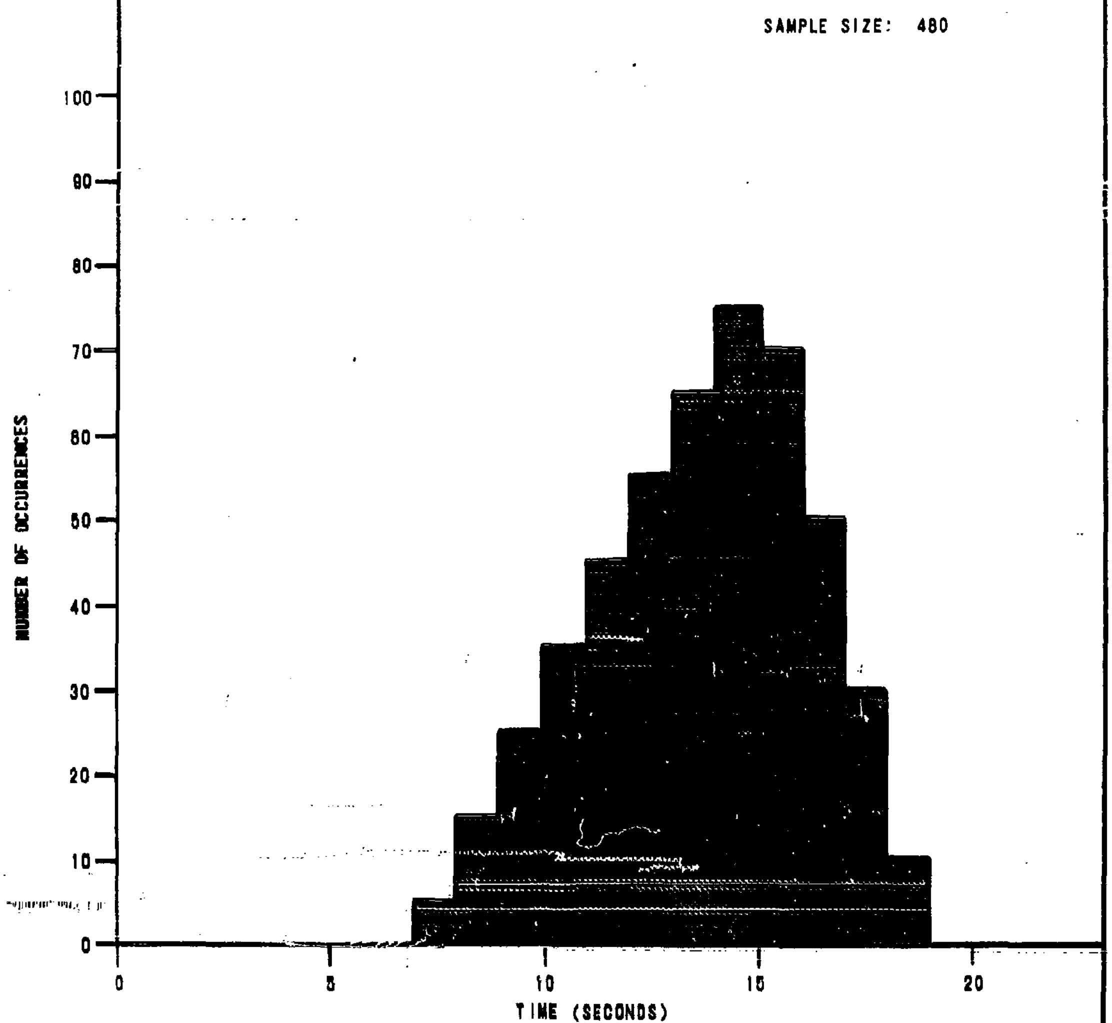
Time Interval (seconds)	Number of occurrences
0-7	0
7-8	5
8-9	15



THE ABOVE HISTOGRAM IS AN EXAMPLE MEANT TO DISPLAY THE DATA HYPOTHETICALLY OBTAINED BY OBSERVING A PARTICULAR EVENT 480 TIMES. THE HISTOGRAM DEPICTS THE NUMBER OF TIMES THE EVENT OCCURRED WITHIN A CERTAIN TIME RANGE. FOR EXAMPLE, 35 TIMES THE EVENT TOOK A TIME BETWEEN 10 AND 11 SECONDS TO COMPLETE.

~~FIG. 7 - APPROXIMATING PDF~~

FIG. 8 - EVENT TIME HISTOGRAM EXAMPLE



THE ABOVE HISTOGRAM IS AN EXAMPLE MEANT TO DISPLAY THE DATA HYPOTHETICALLY OBTAINED BY OBSERVING A PARTICULAR EVENT 480 TIMES. THE HISTOGRAM DEPICTS THE NUMBER OF TIMES THE EVENT OCCURRED WITHIN A CERTAIN TIME RANGE. FOR EXAMPLE, 35 TIMES THE EVENT TOOK A TIME BETWEEN 10 AND 11 SECONDS TO COMPLETE.

FIG. 8 - EVENT TIME HISTOGRAM EXAMPLE

Time Interval (seconds)	Number of occurrences
9-10	25
10-11	35
11-12	45
12-13	55
13-14	65
14-15	75
15-16	70
16-17	50
17-18	30
18-19	10
19 and up	0

From this data a histogram like the one shown in Figure 8 can be constructed. Next an approximating curve composed of line segments is drawn as shown in the overlay for Figure 7. Finally, the end points, which in this case are (7, 0), (15, 80), and (19, 0), are listed. These line segment end points will then be entered into the computer in an appropriate way and will be used to generate random time values obeying this approximating pdf.

The method followed in generating random numbers obeying a distribution characterized by a set of line segments can be seen in the following material. For simplicity first consider a distribution approximated by a single line segment (the extension to N line segments in straight forward and is indicated in the check program shown at the end of this chapter).

The equation for an arbitrary line segment is given by

$$f(x) = A x + B, \; x \epsilon [x_o, x_1]$$

where

$$A = \frac{f(x_1) - f(x_o)}{x_1 - x_o}, \quad B = \frac{x_1 f(x_o) - x_o f(x_1)}{x_1 - x_o}$$

and $x_o < x_1$ (only real-valued, single-valued functions of a real variable are considered). If this function is to be a probability density function (pdf), then for $\int_{-\infty}^{\infty} f(x)\,dx = C$ C should equal 1. The function f can be normalized by

dividing by C. The resulting function, g, is given by $g(x) = \frac{A}{C}x + \frac{B}{C}$. (1)
The cumulative distribution function (cdf) for this function is

$$G(x) = \int_{-\infty}^{x} g(t)\,dt = \int_{x_o}^{x} \left(\frac{A}{C}t + \frac{B}{C}\right)dt.$$

Hence, $G(x) = \frac{1}{C}\left[\frac{A}{2}(x^2 - x_o^2) + B(x - x_o)\right], \quad x \in [x_o, x_1]$ (2)

$$\text{where } C = \int_{-\infty}^{+\infty} f(x)\,dx$$

$$= \int_{x_o}^{x_1} (Ax + B)\,dx$$

$$= \frac{A}{2}(x_1^2 - x_o^2) + B(x_1 - x_o).$$

A well-known theorem* states that if X is a random variable of the continuous type having pdf $g(x)$ and cdf $G(X)$, then the random variable $R = G(X)$ has a uniform distribution with pdf

$$h(r) = 1 \text{ for } o < r < 1$$
$$= 0 \text{ elsewhere.}$$

Hence, it is possible to take a uniformly distributed random variable R and with it arrive at a random variable X having a pdf which is a line segment. This can be done by letting $R = G(x)$ in equation 2 and solving for x.

$$R = \frac{A}{2C}(x^2 - x_o^2) + \frac{B}{C}(x - x_o) = \frac{A}{2C}x^2 + \frac{B}{C}x - \left(\frac{A}{2C}x_o^2 + \frac{B}{C}x_o\right).$$

*See, for example, Hogg and Craig, Introduction to Mathematical Statistics, Macmillan, 1959, p. 157.

Hence,
$$\left(\frac{A}{2C}\right) x^2 + \left(\frac{B}{C}\right) x - \left(\frac{A}{2C} x_o^2 + \frac{B}{C} x_o + R\right) = o.$$

Solving for x yields

$$x = \frac{-B/C \pm \sqrt{B^2/C^2 + 2A/C\left(\frac{A}{2C} x_o^2 + \frac{B}{C} x_o + R\right)}}{A/C}$$

$$= \frac{-B \pm \sqrt{B^2 + A^2 x_o^2 + 2ABx_o + 2ACR}}{A} . \qquad (3)$$

Now let $y_o = f(x_o)$ and $y_1 = f(x_1)$. Then

$$A = \frac{y_1 - y_o}{x_1 - x_o}, \quad B = \frac{x_1 y_o - x_o y_1}{x_1 - x_o},$$

$$C = \frac{A}{2}(x_1^2 - x_o^2) + B(x_1 - x_o)$$

$$= \frac{1}{2}\left(\frac{y_1 - y_o}{x_1 - x_o}\right)(x_1^2 - x_o^2) + \left(\frac{x_1 y_o - x_o y_1}{x_1 - x_o}\right)(x_1 - x_o)$$

$$= \frac{1}{2}\left(\frac{x_1^2 y_1 - x_o^2 y_1 - x_1^2 y_o + x_o^2 y_o}{x_1 - x_o}\right) +$$

$$\frac{x_1^2 y_o - x_o x_1 y_o - x_o x_1 y_1 + x_o^2 y_1}{x_1 - x_o}$$

$$= \frac{1}{2(x_1 - x_o)}(x_1^2 y_1 + x_o^2 y_1 + x_1^2 y_o + x_o^2 y_o - 2x_o x_1 y_o - 2x_o x_1 y_1).$$

Hence,
$$C = \frac{(x_1 - x_o)(y_o + y_1)}{2}$$

$$B^2 = \frac{x_1^2 y_o^2 - x_o^2 y_1^2 - 2x_o x_1 y_o y_1}{(x_1 - x_o)^2}$$

$$A^2 x_o^2 = \frac{x_o^2 y_o^2 + x_o^2 y_1^2 - 2x_o^2 y_o y_1}{(x_1 - x_o)^2}$$

$$2ABx_o = \frac{2x_o x_1 y_o y_1 - 2x_o^2 y_1^2 - 2x_o x_1 y_o^2 + 2x_o^2 y_o y_1}{(x_1 - x_o)^2}$$

$$2ACR = (y_1^2 - y_o^2) R.$$

Substituting in equation 3 yields after simplifying

$$B^2 + A^2 x_o^2 + 2ABx_o + 2ACR = y_o^2 + (y_1^2 - y_o^2) R$$

and hence

$$x = \frac{-(x_1 y_o - x_o y_1) \pm (x_1 - x_o) \sqrt{y_o^2 + (y_1^2 - y_o^2) R}}{y_1 - y_o} \qquad (4)$$

To determine which sign to choose, let $y_1 = y_o + a$ and consider the case where $a \neq o$. Equation (4) now becomes

$$x = \frac{x_o (y_o + a) - x_1 y_o + K (x_1 - x_o) \sqrt{y_o^2 + (y_o^2 + 2ay_o + a^2 - y_o^2) R}}{(y_o + a - y_o)}$$

$$= \frac{1}{a}\left(x_o y_o + x_o a - x_1 y_o + K(x_1 - x_o)\sqrt{y_o^2 + (2ay_o + a^2)R}\right)$$

$$= \frac{1}{a}\left[(x_o - x_1)\,y_o + K\,(x_1 - x_o)\sqrt{y_o^2 + a(2y_o + a)R}\right] + x_o$$

$$\therefore x = x_o + \frac{x_1 - x_o}{a}\left[K\sqrt{y_o^2 + a(2y_o + a)R} - y_o\right]. \qquad (5)$$

When $R = O$, x must equal x_o. Hence let $R = O$ in equation 5.

$$x = x_o + \frac{x_1 - x_o}{a}\left(Ky_o - y_o\right)$$

$$= x_o + \frac{y_o(x_1 - x_o)}{a}(K - 1).$$

Hence K must be plus one.

When $R = 1$, x must equal x_1. Hence let $R = 1$ in equation 5.

$$x = x_o + \frac{x_1 - x_o}{a}\left[K\sqrt{y_o^2 + 2ay_o + a^2} - y_o\right]$$

$$= x_o + \frac{x_1 - x_o}{a}\left[K\ (y_o + a) - y_o\right]$$

$$= x_o + \frac{x_1 - x_o}{a}\left[y_o(K - 1) + Ka\right]$$

Hence $K = +1$ for any $a \neq o$.

The case where $y_o = y_1$, that is, $a = o$, is treated differently. In this case

$$A = \frac{y_1 - y_o}{x_1 - x_o} = O$$

$$B = -x_o\left(\frac{y_1 - y_o}{x_1 - x_o}\right) + y_o = y_o$$

$$C = \frac{A}{2}(x_1^2 - x_o^2) + B(x_1 - x_o) = y_o(x_1 - x_o).$$

Hence

$$G(x) = \frac{B}{C}(x - x_o)$$

$$= \frac{y_o(x - x_o)}{y_o(x_1 - x_o)}$$

$$\therefore G(x) = \frac{x - x_o}{x_1 - x_o}$$

as was expected. Now letting R = G(x) and solving for x it is seen that

$$x = x_o + (x_1 - x_o) R. \qquad (6)$$

In the subroutine that follows, equation 6 is statement 3, while equation 4 is represented by the statements headed by statement 4.

```
      SUBROUTINE RNDNR (N,X,Y,C,T)
      DIMENSION X(10),Y(10),C(10)
      R=RAN(1.)/10.
      DO 1 I=1,N
      IF(R-C(I)) 2,1,1
    1 CONTINUE
    2 W=(R-C(I-1))/(C(I) - C(I-1))
      IF(Y(I)-Y(I-1)) 4,3,4
    3 T=X(I-1) + (X(I)-X(I-1))*W
      RETURN
    4 T=X(I-1)*Y(I)-X(I)*Y(I-1)
      T=T +   (X(I)-X(I-1))*SQRTF(Y(I-1)**2 + (Y(I)**2 - Y(I-1)**2)*W)
      T=T/(Y(I)-Y(I-1))
      RETURN
      END
```

```
C       RANDOM NUMBER GENERATOR CHECK
        DIMENSION X(10),Y(10),A(10),C(10),NN(101)
    100 FORMAT(I2,I3)
    101 FORMAT(20F4.0)
    200 FORMAT(1H ,I2,2XI4)
    201 FORMAT(I4)
      1 READ 100, N,K
        READ 101, (X(I),Y(I), I=1,N)
        A(1)=0
        AT=0
        DO 2 I=2,N
        A(I)=(X(I)-X(I-1))*(Y(I) + Y(I-1))/2.
      2 AT=AT + A(I)
        C(1)=0
        DO 3 I=2,N
      3 C(I)=C(I-1) + A(I)/AT
        Z=K
        Z=(X(N) - X(1))/Z
        L=K + 1
        DO 4 I=1,L
      4 NN(I)=0
        DO 6 I=1,1000
        IF(SENSE SWITCH 1)8,9
      8 TYPE 201,I
      9 CALL RNDNR (N,X,Y,C,T)
        DO 5 J=1,L
        B=J-1
        IF(T-(X(1) + Z*B)) 6,5,5
      5 CONTINUE
      6 NN(J)=NN(J) + 1
        DO 7 I=1,L
      7 PRINT 200,I,NN(I)
        PAUSE
        GO TO 1
        END
```

III. DATA ANALYSIS

The measure to be used for evaluating information retrieval systems at present is response time, which has been defined as the time it takes for the system to respond to a given request. The purpose of the simulation model is to provide representative system response time data. Based on this data an evaluator would want to know at least two things about the system:

1. whether or not the system is acceptable according to a given time constraint.
2. whether or not the system can be improved to make it acceptable, or better.

To date the efforts of this task have been restricted to the first question. There are two problems associated with question one: (1) It might be desirable to know in what response time T an evaluator can have a P percent confidence for the given system and (2) the evaluator would want to know whether he should accept or reject the system if he must be P percent confident that the response time T is less than some required time T*. These two problems can be answered by means of the two programs which follow

1. DATA REDUCTION PROGRAM

The general process followed for data reduction is shown in Figure 9. The purpose of the data reduction program is to develop a histogram (response time vs. frequency of occurrence) and pertinent statistics by examining the simulation output cards. This output can be examined in several ways, some of which are as follows.

a. Obtain a complete summary analysis of the simulation output. In this case all output cards are examined and they produce one summary listing.

b. Obtain partial summaries; that is, the data reduction program would examine for example, the first 100 iterations and produce a summary and then examine the next 100 and produce a composite summary of the entire 200 iterations, and so on

c. The time histogram developed can be varied by considering different time interval subdivisions.

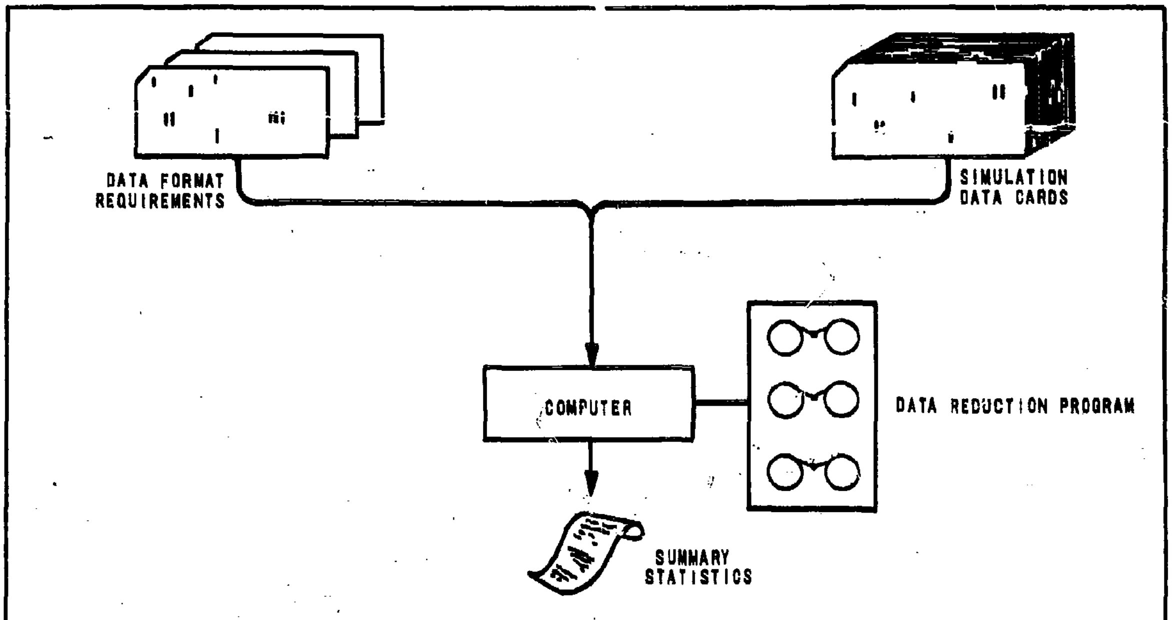


FIG. 9 - DATA REDUCTION PROCESS

Figure 10 is the data reduction program flow chart. As will be seen in the program listing which follows, D1 and D2 represent all of the data on two separate output cards.

The program requires some preliminary data prior to accepting the simulation output data. This preliminary data (data format requirements) are on three cards. The first card has the number of iterations (simulation output cards) to be examined and the desired number of summaries (b. above). The second card contains basic data for the histogram: number of time intervals desired, the length of each interval (which is the same for all intervals), and the initial time. The third card has data about the number of query types, the maximum number of queries, and the number of input and output devices involved in the simulation. The simulation output cards follow these three cards, and a blank card follows the output cards (another program requirement at present).

The output of the data reduction program consists of three pages of data (illustrated in section 2). On the first page (labeled page 1), which is predominantly time data, will be found the following information:

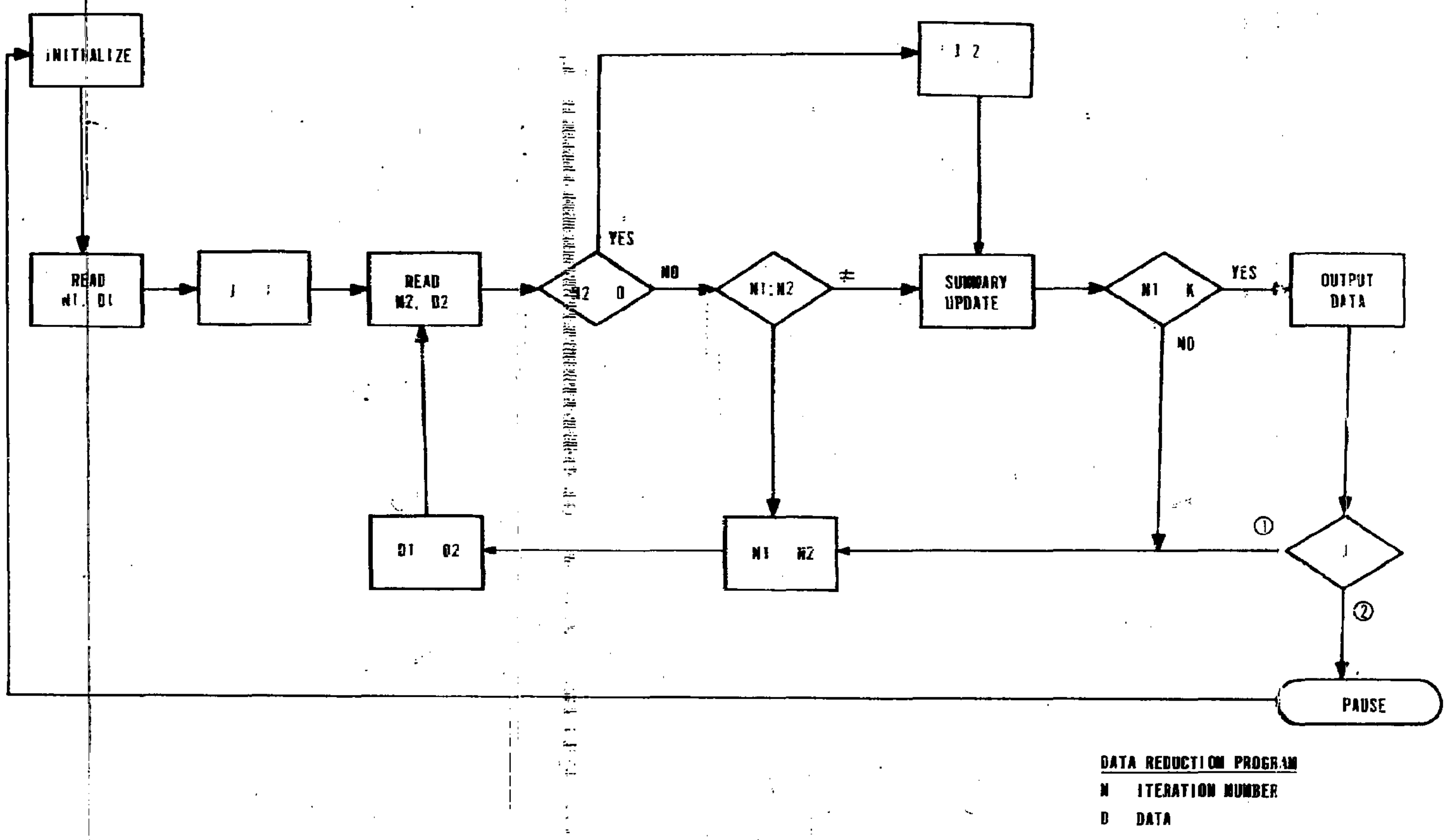


FIG 10 - DATA REDUCTION PROGRAM FLOW CHART

1. The number of iterations considered;
2. The average time per iteration;
3. The time variance;
4. The time standard deviation;
5. Time histogram data (response time vs. number of occurrences);
6. Time histogram data (response time vs. frequency of occurrence);
7. Time histogram data (response time vs. cumulative frequency of occurrence).

The second page (labeled page 2), which is concerned with query data, contains the following data:

1. Number of queries generated in the previously specified number of iterations;
2. Average number of queries per iteration;
3. Query occurrence per iteration data;
4. Query type occurrence data.

The last page (labeled page 3) consists of information about the utilization of the I/O devices.

```
C       DATA REDUCTION PROGRAM FOR MOD 2
        DIMENSION NQS(11),NT(25),F(25),NTY(10),NTD(10),
       1NTT1(10),NTT2(10),ND1(10),ND2(10)
    100 FORMAT(2I5)
    101 FORMAT(3I5)
    102 FORMAT(4I3)
    103 FORMAT(I4,F10.2,I4,10I3,10I3)
    200 FORMAT(1H1,39HINFORMATION RETRIEVAL SYSTEM SIMULATION,13X,8H(PAGE
       11))
    201 FORMAT(1H0,23HNUMBER OF ITERATIONS = ,I5)
    202 FORMAT(1H0,9HTIME DATA)
    203 FORMAT(1H ,5X,29HAVERAGE TIME PER ITERATION = ,F10.2,1X,7HSECONDS)
    204 FORMAT(1H ,5X,11HVARIANCE = ,18X,F10.3)
    205 FORMAT(1H0,18HTIME INTERVAL DATA)
    206 FORMAT(1H0,5X,8HINTERVAL,5X,21HNUMBER OF OCCURRENCES,5X,
       19HFREQUENCY,4X,15HCUMULATIVE PROB)
    207 FORMAT(1H ,5X,I4,1H-,I4,10X,I5,16X,F6.3,10X,F6.3)
    208 FORMAT(1H ,5X,I4,12H AND GREATER,3X,I5,16X,F6.3)
    209 FORMAT(1H1,39HINFORMATION RETRIEVAL SYSTEM SIMULATION,13X,8H(PAGE
       12))
    210 FORMAT(1H0,10HQUERY DATA)
    211 FORMAT(1H ,5X,19HNUMBER OF QUERIES =,I5)
    212 FORMAT(1H ,5X,41HAVERAGE NUMBER OF QUERIES PER ITERATION =,F6.2)
    213 FORMAT(1H0,17HQUERIES/ITERATION,3X,21HNUMBER OF OCCURRENCES,3X,17H
       1AVERAGE/ITERATION)
    214 FORMAT(1H ,8X,I2,17X,I5,16X,F6.2)
    215 FORMAT(1H ,3X,9HMORE THAN,I3,12X,I5,16X,F6.2)
    216 FORMAT(1H0,10HQUERY TYPE,10X,21HNUMBER OF OCCURRENCES,3X,13HAVERAG
       1E/QUERY)
    217 FORMAT(1H ,8X,I2,17X,I5,15X,F6.2)
    218 FORMAT(1H1,39HINFORMATION RETRIEVAL SYSTEM SIMULATION,13X,8H(PAGE
       13))
    219 FORMAT(1H0,24HINPUT/OUTPUT DEVICE DATA)
    220 FORMAT(1H0,14HUTILIZATION OF,10X,17HAVERAGE USE/QUERY,3X,21HAVERAG
       1E USE/ITERATION)
    221 FORMAT(1H ,12HINPUT DEVICE,I3,1H=,I4,9X,F6.2,17X,F6.2)
    222 FORMAT(1H ,13HOUTPUT DEVICE,I2,1H=,I4,9X,F6.2,17X,F6.2)
    223 FORMAT(1H ,5X,20HSTANDARD DEVIATION =,9X,F10.3)
      1 READ 100,NR,NN
        L=NR/NN
        READ101,NTI,LI,IS
        READ 102,M,NI,NO,NQ
        N=2*NQ+1
        K=NI+NO
        SS=0
        SSQ=0
        NTQ=0
        DO 2 I=1,N
      2 NQS(I)=0
        DO 3 I=1,NTI
```

```
    NT(I)=0
  3 F(I)=IS+LI*I
    DO 4 I=1,M
  4 NTY(I)=0
    DO 5 I=1,K
  5 NTD(I)=0
    READ 103,N1,X,NQ1,(NTT1(I),I=1,M),(ND1(I),I=1,K)
    J=1
  6 READ 103,N2,Y,NQ2,(NTT2(I),I=1,M),(ND2(I),I=1,K)
    IF(N2)7,11,7
  7 IF(N1-N2)12,8,12
  8 N1=N2
    X=Y
    DO 9 I=1,M
  9 NTT1(I)=NTT2(I)
    DO 10 I=1,K
 10 ND1(I)=ND2(I)
    NQ1=NQ2
    GO TO 6
 11 J=2
 12 SS=SS+X
    SSQ=SSQ+X**2
    DO 13 I=1,N
    IF(NQ1-I)13,14,13
 13 CONTINUE
 14 NQS(I)=NQS(I)+1
    NTQ=NTQ+NQ1
    DO 15 I=1,NTI
    IF(X-F(I))16,15,15
 15 CONTINUE
 16 NT(I)=NT(I)+1
    DO 17 I=1,M
 17 NTY(I)=NTY(I)+NTT1(I)
    DO 18 I=1,K
 18 NTD(I)=NTD(I)+ND1(I)
    DO 19 I=1,NN
    IF(N1-I*L)8,20,19
 19 CONTINUE
 20 W=N1
    XB=SS/W
    XV=SSQ/W-XB**2
    XS=SQRTF(XV)
    PRINT 200
    PRINT 201,N1
    PRINT 202
    PRINT 203,XB
    PRINT 204,XV
    PRINT 223,XS
    PRINT 205
    PRINT 206
```

```
      II=0
      FK=NT(1)
      FK=FK/W
      CP=FK
      PRINT 207,II,F(1),NT(1),FK,CP
      II=NTI-1
      DO 21 I=2,II
      MM=I-1
      FK=NT(I)
      FK=FK/W
      CP=CP + FK
   21 PRINT 207,F(MM),F(I),NT(I),FK,CP
      FK=NT(NTI)
      FK=FK/W
      PRINT 208,F(II),NT(NTI),FK
      PRINT 209
      PRINT 210
      PRINT 211,NTQ
      G=NTQ
      G1=G/W
      PRINT 212,G1
      PRINT 213
      II=N-1
      DO 22 I=1,II
      H=NQS(I)
      H=H/W
   22 PRINT 214,I,NQS(I),H
      H=NQS(N)
      H=H/W
      PRINT 215,II,NQS(N),H
      PRINT 216
      DO 23 I=1,M
      H=NTY(I)
      H=H/G
   23 PRINT 217,I,NTY(I),H
      PRINT 218
      PRINT 219
      PRINT 220
      DO 24I=1,NI
      G1=NTD(I)
      G2=G1/G
      G3=G1/W
   24 PRINT 221,I,NTD(I),G2,G3
      DO 25 I=1,NO
      K1=NI+I
      G1=NTD(K1)
      G2=G1/G
      G3=G1/W
   25 PRINT 222,I,NTD(K1),G2,G3
      GO TO (8,26),J
```

2. EXAMPLE - DATA REDUCTION PROGRAM OUTPUT

The form of the data reduction program output can be illustrated by considering the numerical output data given in Appendix B. There were 500 samples generated in this particular run. The three required header cards for the data reduction program are

CARD	VALUES
1	NR = 500, NN = 5
2	NTI = 10, LI = 50, IS = 150
3	M = 5, NI = 2, NO = 3, NQ = 3

As previously mentioned, the header cards are followed by the iteration cards, the last of which is followed by a blank card. The following pages indicate the output produced by the data reduction program. For convenience, histograms have been prepared (Figures 11 - 15) and follow the output example. It should be mentioned again that this is a numerical example used to show the form and substance of the program output and meant to illustrate the capabilities of the program.

Using the example cited above, assume that someone wants to know in what response time of the system he can be, say, 90% confident. An examination of Figure 15 (b) shows that the system response time will be less than 450 seconds about 90% of the time. Hence, based on this data he can expect that 90% of the time he will get a response to a request from this system in something less than seven and a half minutes.

As can be seen in the program listing, there is some degree of freedom allowed in the output format. For example, there can be up to 25 time subdivisions if desired. The output, in its present form, permits an examination of page 1 for time data, page 2 for query data, and page 3 for I/O device data.

INFORMATION RETRIEVAL SYSTEM SIMULATION (PAGE 1)

NUMBER OF ITERATIONS = 100

TIME DATA

AVERAGE TIME PER ITERATION = 327.83 SECONDS
VARIANCE = 7300.010
STANDARD DEVIATION = 85.440

TIME INTERVAL DATA

INTERVAL	NUMBER OF OCCURRENCES	FREQUENCY	CUMULATIVE PROB
0- 150	0	0.000	0.000
150- 200	6	.060	.060
200- 250	12	.120	.180
250- 300	23	.230	.410
300- 350	21	.210	.620
350- 400	19	.190	.810
400- 450	8	.080	.890
450- 500	8	.080	.970
500- 550	3	.030	1.000
550 AND GREATER	0	0.000	

QUERY DATA

NUMBER OF QUERIES = 163

AVERAGE NUMBER OF QUERIES PER ITERATION = 1.63

QUERIES/ITERATION	NUMBER OF OCCURRENCES	AVERAGE/ITERATION
1	52	.52
2	33	.33
3	15	.15
4	0	0.00
5	0	0.00
6	0	0.00
MORE THAN 6	0	0.00

QUERY TYPE	NUMBER OF OCCURRENCES	AVERAGE/QUERY
1	71	.43
2	50	.30
3	24	.14
4	11	.06
5	7	.04

INFORMATION RETRIEVAL SYSTEM SIMULATION (PAGE 3)

INPUT/OUTPUT DEVICE DATA

UTILIZATION OF	AVERAGE USE/QUERY	AVERAGE USE/ITERATION
INPUT DEVICE 1= 124	.76	1.24
INPUT DEVICE 2= 39	.23	.39
OUTPUT DEVICE 1= 100	.61	1.00
OUTPUT DEVICE 2= 25	.15	.25
OUTPUT DEVICE 3= 38	.23	.38

INFORMATION RETRIEVAL SYSTEM SIMULATION (PAGE 1)

NUMBER OF ITERATIONS = 200

TIME DATA

AVERAGE TIME PER ITERATION = 325.89 SECONDS
VARIANCE = 7419.110
STANDARD DEVIATION = 86.134

TIME INTERVAL DATA

INTERVAL	NUMBER OF OCCURRENCES	FREQUENCY	CUMULATIVE PROB
0- 150	0	0.000	0.000
150- 200	11	.055	.055
200- 250	29	.145	.200
250- 300	48	.240	.440
300- 350	40	.200	.640
350- 400	31	.155	.795
400- 450	18	.090	.885
450- 500	16	.080	.965
500- 550	7	.035	1.000
550 AND GREATER	0	0.000	

QUERY DATA

NUMBER OF QUERIES = 333

AVERAGE NUMBER OF QUERIES PER ITERATION = 1.66

QUERIES/ITERATION	NUMBER OF OCCURRENCES	AVERAGE/ITERATION
1	101	.50
2	65	.32
3	34	.17
4	0	0.00
5	0	0.00
6	0	0.00
MORE THAN 6	0	0.00

QUERY TYPE	NUMBER OF OCCURRENCES	AVERAGE/QUERY
1	128	.38
2	99	.29
3	56	.16
4	35	.10
5	15	.04

INPUT/OUTPUT DEVICE DATA

UTILIZATION OF	AVERAGE USE/QUERY	AVERAGE USE/ITERATION
INPUT DEVICE 1= 238	.71	1.19
INPUT DEVICE 2= 95	.28	.47
OUTPUT DEVICE 1= 201	.60	1.00
OUTPUT DEVICE 2= 39	.11	.19
OUTPUT DEVICE 3= 93	.27	.46

INFORMATION RETRIEVAL SYSTEM SIMULATION (PAGE 1)

NUMBER OF ITERATIONS = 300

TIME DATA

AVERAGE TIME PER ITERATION = 327.86 SECONDS
VARIANCE = 7768.240
STANDARD DEVIATION = 88.137

TIME INTERVAL DATA

INTERVAL	NUMBER OF OCCURRENCES	FREQUENCY	CUMULATIVE PROB
0- 150	0	0.000	0.000
150- 200	19	.063	.063
200- 250	42	.140	.203
250- 300	68	.226	.429
300- 350	60	.200	.629
350- 400	47	.156	.786
400- 450	30	.100	.886
450- 500	23	.076	.963
500- 550	10	.033	.996
550 AND GREATER	1	.003	

QUERY DATA

NUMBER OF QUERIES = 506

AVERAGE NUMBER OF QUERIES PER ITERATION = 1.68

QUERIES/ITERATION	NUMBER OF OCCURRENCES	AVERAGE/ITERATION
[illegible]	153	.51
2	88	.29
3	59	.19
4	0	0.00
5	0	0.00
6	0	0.00
MORE THAN 6	0	0.00

QUERY TYPE	NUMBER OF OCCURRENCES	AVERAGE/QUERY
1	197	.38
2	164	.32
3	74	.14
4	45	.08
5	26	.05

INPUT/OUTPUT DEVICE DATA

UTILIZATION OF	AVERAGE USE/QUERY	AVERAGE USE/ITERATION
INPUT DEVICE 1= 365	.72	1.21
INPUT DEVICE 2= 141	.27	.47
OUTPUT DEVICE 1= 309	.61	1.03
OUTPUT DEVICE 2= 58	.11	.19
OUTPUT DEVICE 3= 139	.27	.46

INFORMATION RETRIEVAL SYSTEM SIMULATION (PAGE 1)

NUMBER OF ITERATIONS = 400

TIME DATA

AVERAGE TIME PER ITERATION = 329.05 SECONDS

VARIANCE = 7357.300

STANDARD DEVIATION = 85.774

TIME INTERVAL DATA

INTERVAL	NUMBER OF OCCURRENCES	FREQUENCY	CUMULATIVE PROB
0- 150	1	.002	.002
150- 200	22	.055	.057
200- 250	52	.130	.187
250- 300	91	.227	.415
300- 350	83	.207	.622
350- 400	68	.170	.792
400- 450	41	.102	.895
450- 500	28	.070	.965
500- 550	13	.032	.997
550 AND GREATER	1	.002	

QUERY DATA

NUMBER OF QUERIES = 676

AVERAGE NUMBER OF QUERIES PER ITERATION = 1.69

QUERIES/ITERATION	NUMBER OF OCCURRENCES	AVERAGE/ITERATION
1	201	.50
2	122	.30
3	77	.19
4	0	0.00
5	0	0.00
6	0	0.00
MORE THAN 6	0	0.00

QUERY TYPE	NUMBER OF OCCURRENCES	AVERAGE/QUERY
1	259	.38
2	217	.32
3	99	.14
4	67	.09
5	34	.05

INPUT/OUTPUT DEVICE DATA

UTILIZATION OF	AVERAGE USE/QUERY	AVERAGE USE/ITERATION
INPUT DEVICE 1= 482	.71	1.20
INPUT DEVICE 2= 194	.28	.48
OUTPUT DEVICE 1= 401	.59	1.00
OUTPUT DEVICE 2= 85	.12	.21
OUTPUT DEVICE 3= 190	.28	.47

INFORMATION RETRIEVAL SYSTEM SIMULATION (PAGE 1)

NUMBER OF ITERATIONS = 500

TIME DATA

AVERAGE TIME PER ITERATION = 330.50 SECONDS
VARIANCE = 7618.840
STANDARD DEVIATION = 87.285

TIME INTERVAL DATA

INTERVAL	NUMBER OF OCCURRENCES	FREQUENCY	CUMULATIVE PROB
0- 150	1	.002	.002
150- 200	26	.052	.054
200- 250	67	.134	.188
250- 300	111	.222	.410
300- 350	101	.202	.612
350- 400	89	.178	.790
400- 450	51	.102	.892
450- 500	33	.066	.958
500- 550	18	.036	.994
550 AND GREATER	3	.006	

QUERY DATA

NUMBER OF QUERIES = 844

AVERAGE NUMBER OF QUERIES PER ITERATION = 1.68

QUERIES/ITERATION	NUMBER OF OCCURRENCES	AVERAGE/ITERATION
1	251	.50
2	154	.30
3	95	.19
4	0	0.00
5	0	0.00
6	0	0.00
MORE THAN 6	0	0.00

QUERY TYPE	NUMBER OF OCCURRENCES	AVERAGE/QUERY
1	326	.38
2	272	.32
3	121	.14
4	81	.09
5	44	.05

INFORMATION RETRIEVAL SYSTEM SIMULATION (PAGE 3)

INPUT/OUTPUT DEVICE DATA

UTILIZATION OF	AVERAGE USE/QUERY	AVERAGE USE/ITERATION
INPUT DEVICE 1= 605	.71	1.21
INPUT DEVICE 2= 239	.28	.47
OUTPUT DEVICE 1= 500	.59	1.00
OUTPUT DEVICE 2= 105	.12	.21
OUTPUT DEVICE 3= 239	.28	.47

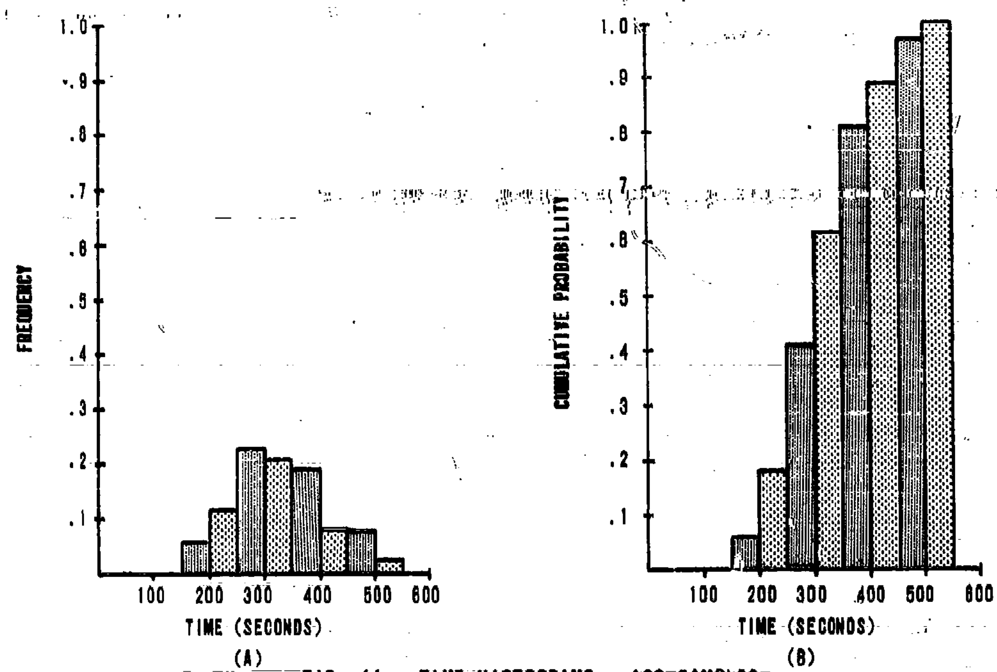


FIG. 11 - TIME HISTOGRAMS - 100 SAMPLES

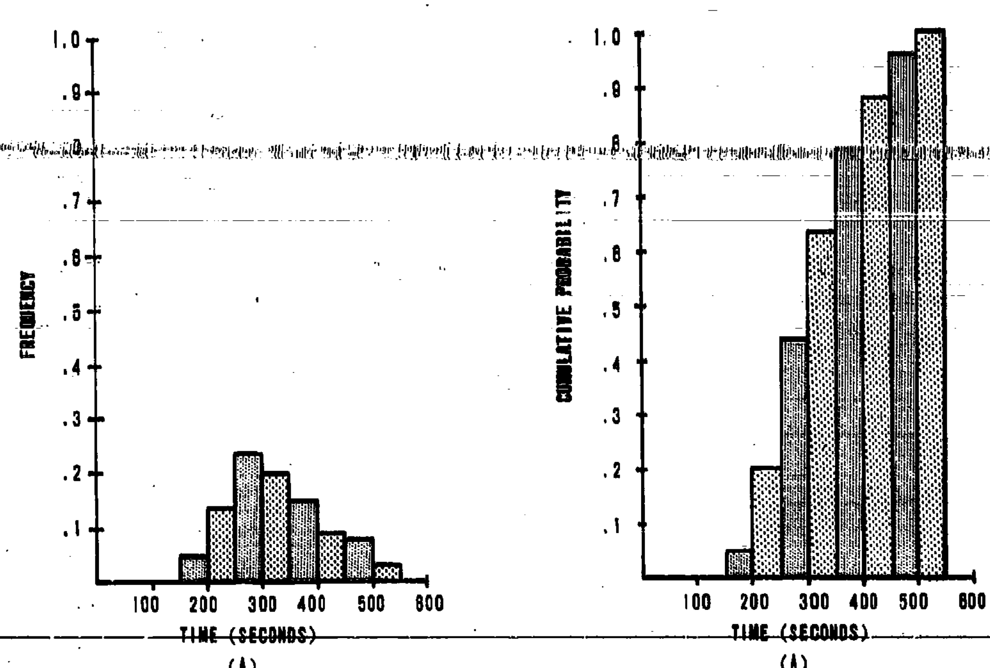


FIG. 12 - TIME HISTOGRAMS - 200 SAMPLES

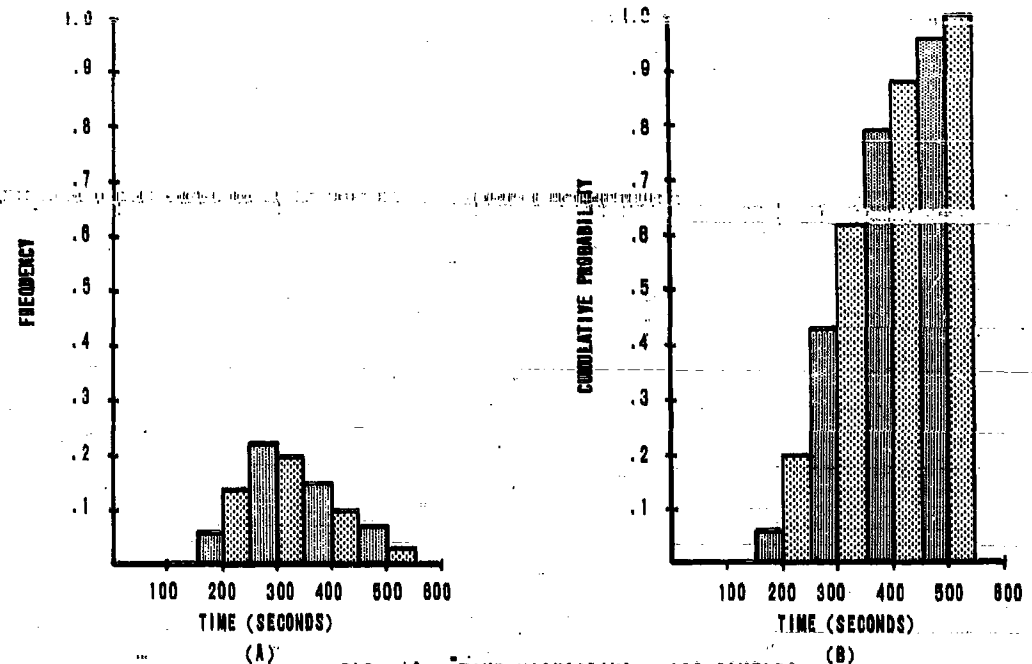


FIG. 13 - TIME HISTOGRAMS - 300 SAMPLES

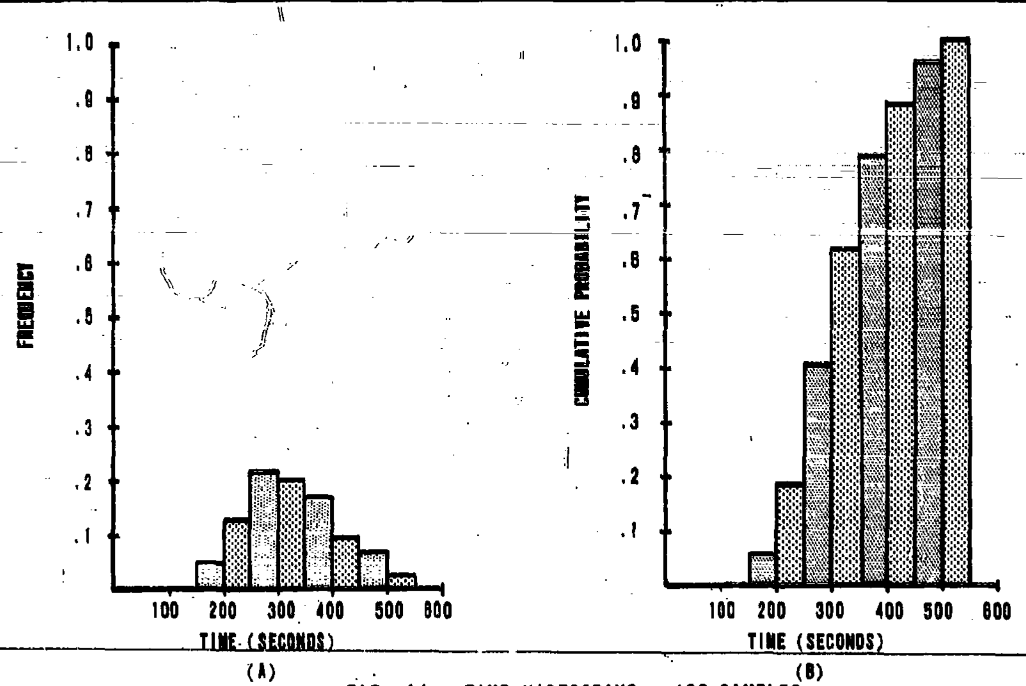


FIG. 14 - TIME HISTOGRAMS - 400 SAMPLES

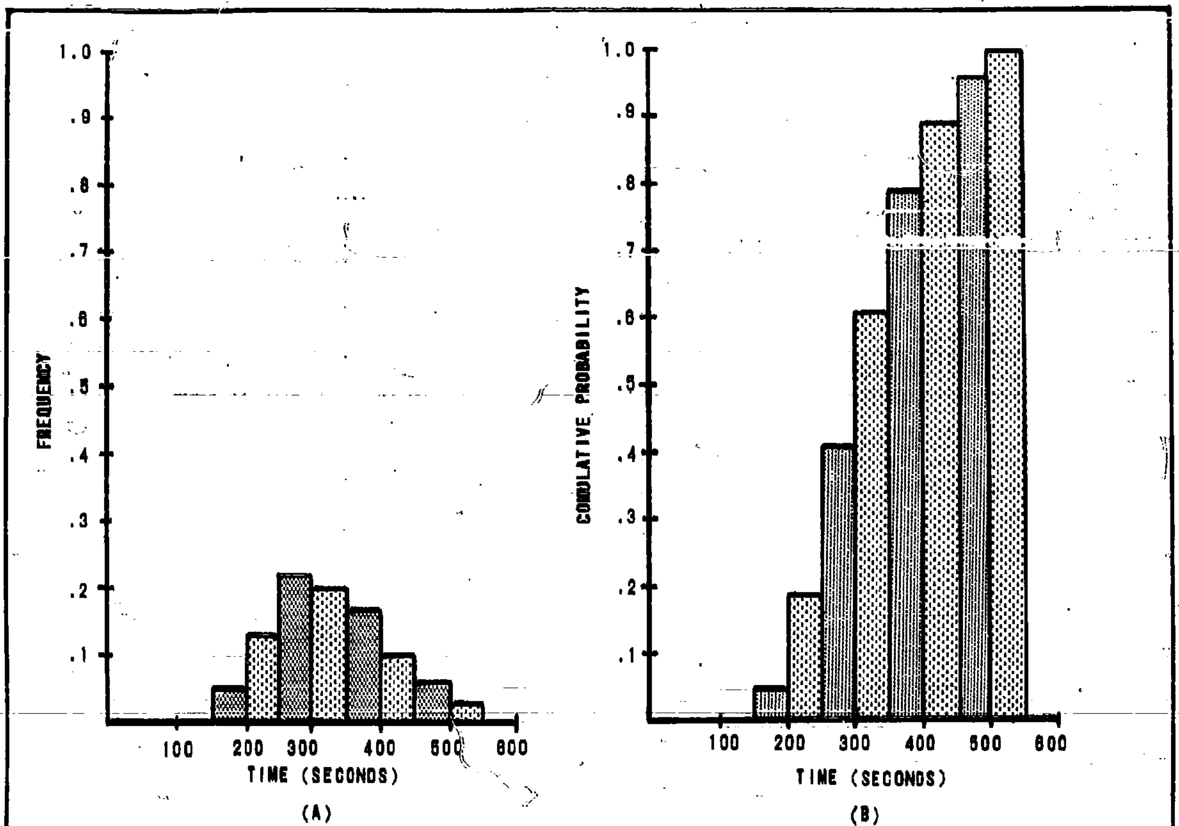


FIG. 15 - TIME HISTOGRAMS - 500 SAMPLES

3. ANALYSIS PROGRAM

The analysis program is a user's problem solver. Fundamentally there are only two questions the system evaluator would want to have answered:

1. Given a response time T, he would want to know the probability that the system could respond in this time.
2. Given a system acceptance criterion, he would want to know with P percent confidence that the response time of the system was less than some specified time T.

The analysis program, just as the data reduction program, examines the simulation output cards (see Figure 16). To do this, two header cards are necessary. The first header card has data about T and **P** (as defined above).

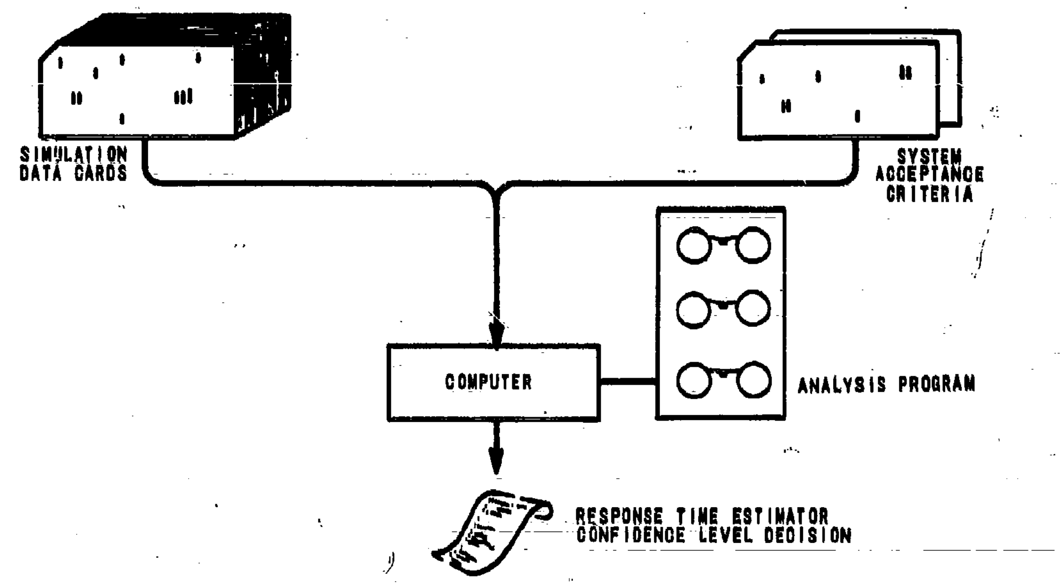


FIG. 16 - ANALYSIS PROCESS

If P is equal to zero, the program interprets this as meaning question one (above) is to be answered; otherwise (i.e., $P \neq 0$) question two. The second header card has on it the number of iterations, N, to be examined. The header cards are followed by N iteration cards, the last of which is followed by a blank card. The procedure followed by the program is shown in Figure 17. Specific input formats appear in the program listing.

The output for question one (given T, find P) is simply the printed statement

PROBABILITY TIME LESS THAN (T) SEC. IS (P),

where numerical values replace T and P. For example, consider the simulation output data of Appendix B. Suppose the evaluator wants to know the confidence level P for a response time of 500 seconds. Since this is question one, the first

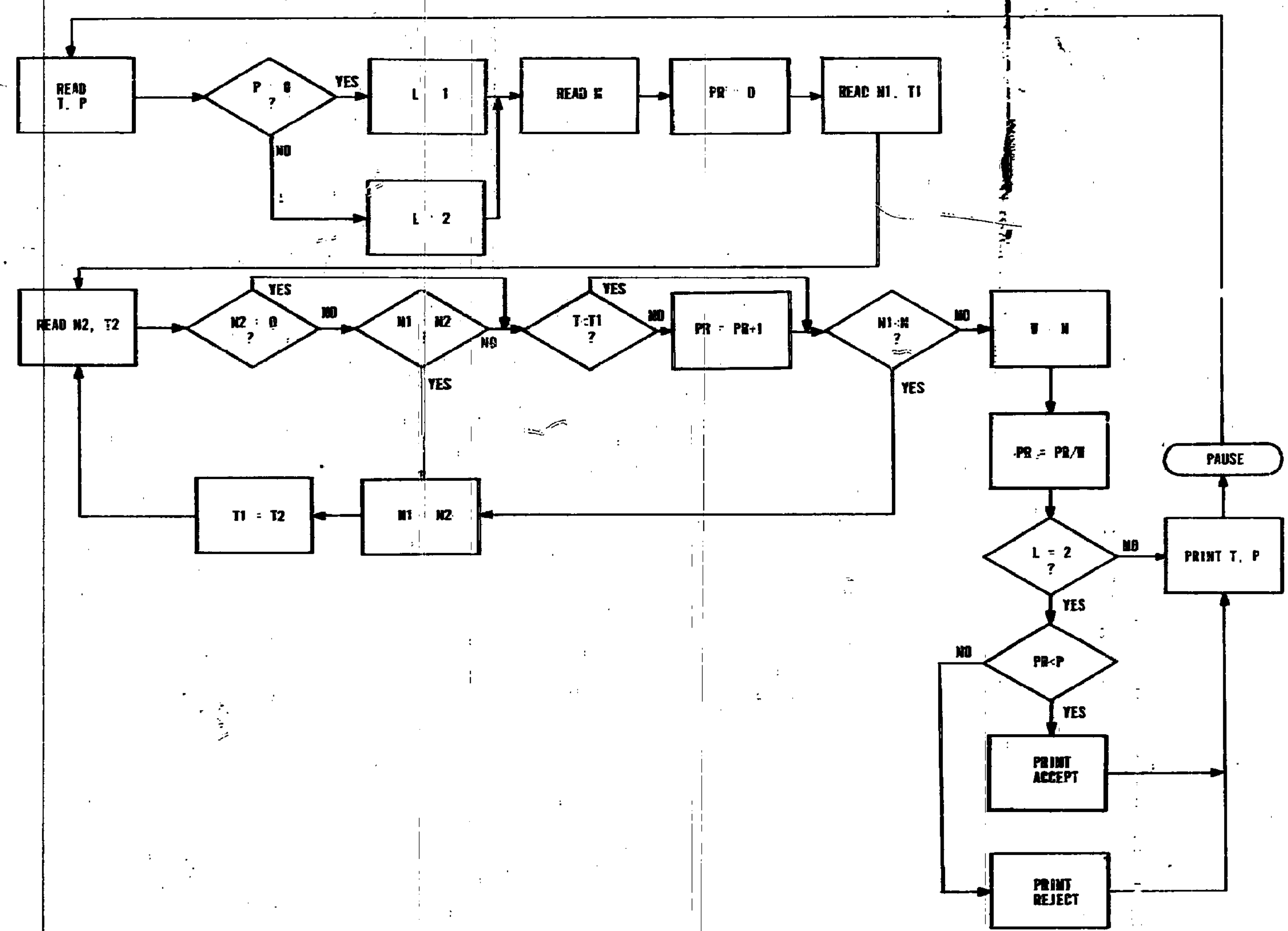


FIG. 17 - ANALYSIS PROGRAM FLOW CHART

header card would have T = 500 and P = O. The second header card would have the number of iterations to be examined, and in this case N = 500. These header cards are followed by the simulation output cards (again, these are followed by a blank card). After the cards have been read by the computer, an answer would be printed out which would be

PROBABILITY TIME LESS THAN 500.00 SEC. IS .958

This answer is interpreted as meaning that 95.8% of the time this system would respond to a given request for information in less than 500 seconds.

The output of question two consists of two statements

(X) SYSTEM SINCE

PROBABILITY TIME LESS THAN (T) SEC. IS (P),

where (X) will be either ACCEPT or REJECT, and there will be numerical values, as with question one, for T and P. Suppose it is required that for an acceptable system the response time must be less than 500 seconds with a confidence level of 90%. The printed output answer would be

ACCEPT SYSTEM SINCE

PROBABILITY TIME LESS THAN 500.00 SEC. IS .958

As a second example, suppose it is necessary to be sure that 90% of the time the response time is less than 400 seconds. In this case (for the same data) the printed output would be

REJECT SYSTEM SINCE

PROBABILITY TIME LESS THAN 400.00 SEC. IS .790.

Marginal rejection cases, such as reject because the calculated P value is just slightly less than the required confidence level, should be interpreted by the user to determine the system's acceptance or rejection.

```
C       ANALYSIS PROGRAM
  100 FORMAT(F8.2,2X,F6.3)
  101 FORMAT(I5)
  102 FORMAT(I4,F10.2)
  200 FORMAT(36HENTER SIMULATION CARDS - PRESS START)
  201 FORMAT(1H0,26HPROBABILITY TIME LESS THAN,F8.2,2X,4HSEC.,3H IS,
     1F6.3)
  202 FORMAT(1H0,19HACCEPT SYSTEM SINCE)
  203 FORMAT(1H0,19HREJECT SYSTEM SINCE)
 1000 READ 100,T,P
      IF(P)2,1,2
    1 L=1
      GO TO 3
    2 L=2
    3 TYPE 200
      PAUSE
      READ 101,N
      PR=0
      READ 102,N1,T1
    4 READ 102,N2,T2
      IF(N2)5,7,5
    5 IF(N1-N2)7,6,7
    6 N1=N2
      T1=T2
      GO TO 4
    7 IF(T-T1)9,8,8
    8 PR=PR+1.
    9 IF(N1-N)6,10,10
   10 W=N
      PR=PR/W
      IF(L-1)12,11,12
   11 PRINT 201,T,PR
      PAUSE
      GO TO 1000
   12 IF(PR-P)14,13,13
   13 PRINT 202
      GO TO 11
   14 PRINT 203
      GO TO 11
      END
```

IV. SUMMARY AND RECOMMENDATIONS

To date, our primary goal has been the development of a basis for evaluating information retrieval systems, based on the time it takes for the system to respond to a given request for information. Recent discussions of the model substantiate the choice and use of response time as a primary measure for simulating and evaluating retrieval systems. The model is still in the early stages of development. Its logic has been developed along with two variations in the model (each of which has been programmed and run on our IBM 1620 computer). The evaluation aspect of the problem has been examined and two programs have been developed which use the simulation model's output. The first of these is a data reduction program which produces a summary of pertinent statistics obtained from the simulation run. The second is an analysis program which can be considered a model in the sense that once the user has decided on the acceptance level of the system, the model can examine the simulated data and determine whether or not the given system is accepted by the specified standards.

In the coming year the existing models will be refined and extended by including more equipment characteristics (that is, by introducing various equipment parameters such as card read rate, tape density, word size, etc., to be used in the simulation) as well as including query, record and output characteristics. This work should complete the simulation model of the computer-based information retrieval system.

After the model has been completed, a model of a manual information retrieval system will be developed, in which the data may be stored as hard copy and retrieved entirely by human operation. Equipment characteristics to be included in this case are such factors as storage capacity and access time. Other factors which might be included are cataloguing procedures (what goes where), query variations (ways of asking for the same thing), and so on. When developed, the model will be programmed and run on HRB-Singer's IBM 1620 computer. Modifications to the data reduction and analysis programs will be made where necessary.

Once the manual system model has been developed, it will be integrated with the computer-based retrieval model into what could be called a General Information Retrieval System Simulation (GIRSS) model, or G model. It would be this model which would find the greatest application since most information

retrieval systems are a composite of what has been called manual and computer systems. The G model could be extended to the more general data processing systems or it could be tailored to a specific system, as shown in Figure 18. This would complete the major work on information retrieval system evaluation using the response time measure.

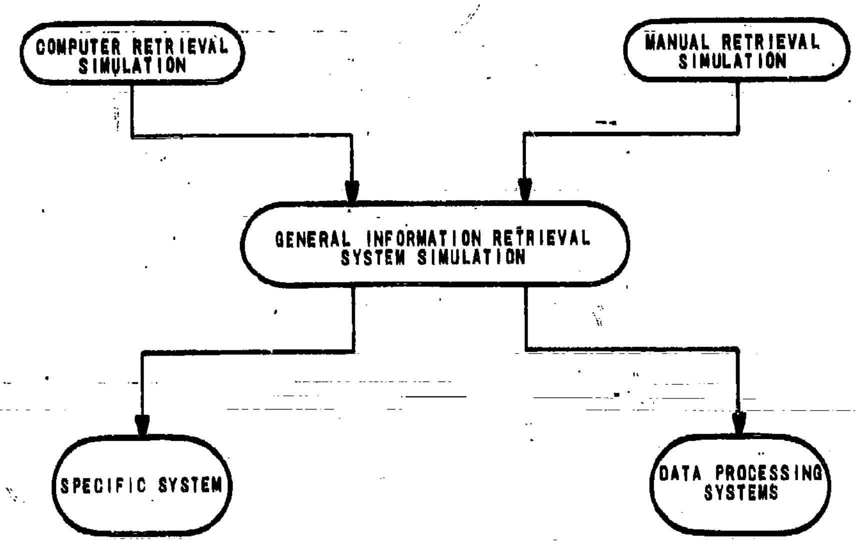


FIG. 18 - MODEL DEVELOPMENT AND APPLICATION

It is a rare system today that can be judged solely on a single criterion. If there were only one system which could do the retrieval job in the required time, then a user would have little choice in deciding which system he should purchase. However, present-day technology allows the user to choose and tailor the procedures and components of his system as he sees fit -- in fact, the problem now is which procedures and which equipment to choose. In a sense, the user has an allocation problem; he has a given amount of resources (cash, time available for a response, space) and he must choose the best possible fit of equipment and procedures to make up his system. The present and future work will aid the user (or manager) in these decisions. The manager can use the G model, for example, to determine whether or not the system he has selected will satisfy his time requirements; but this is not the only requirement

he has. Models of the system using other evaluation (decision) measures will also be of considerable value to him. It is these measures and models that this task will consider next. Our fundamental procedure is diagrammed in Figure 19.

A future goal of this task could be the development of an information retrieval system model which would be based on those measures the user must consider. The user would specify, for example, his cost constraints, response

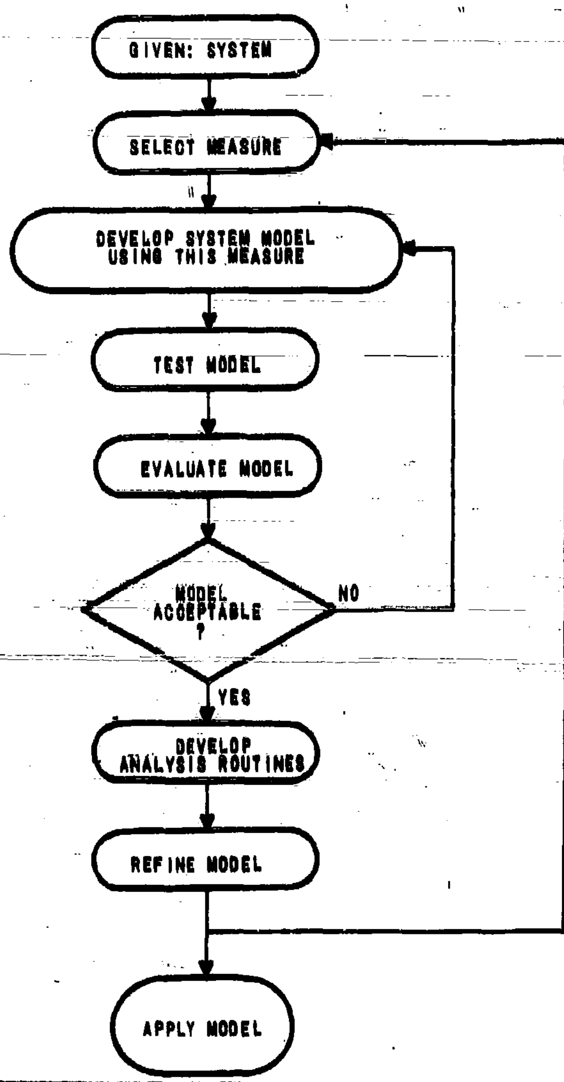


FIG. 19 - ITERATIVE PROCEDURE FOR THE SIMULATION AND EVALUATION OF INFORMATION RETRIEVAL SYSTEMS

time constraints, space restrictions, data volume considerations and enter these data into the model. Also available for entry would be state-of-the-art equipment characteristics and normal event sequences for retrieval systems, along with their corresponding time distributions, cost/operation, and other data. The model could then select and recommend the best (i.e., optimum in some sense) fit under the given constraints. Whether or not this goal is attainable at present, it does provide both motivation and direction for the work to be performed by this task in the future.

APPENDIX A

INPUT EXAMPLE

The data given in this section shows the numerical form of the input for Mod II based on the example described in section 2, Chapter II. This input is composed of simulation data and system data. At present the simulation is simply the number of iterations to be considered. The system data is a numerical description of the system being simulated.

TABLE 4. INFORMATION RETRIEVAL SYSTEM SIMULATION EXAMPLE - INPUT DATA FOR MOD II							
Variable	Value	Card Column	Card Number	Variable	Value	Card Column	Card Number
NR	500	3-5	1	PT (4)	.1	19, 20	14
M	5	3	2	PT (5)	.05	25-27	14
NI	2	6	2	PE (1, 1)	.9	1, 2	15
NO	3	9	2	PE (1, 2)	.1	7, 8	15
NQ	3	12	2	PE (2, 1)	.8	1, 2	16
T(1, 1)	.76	1-3	3	PE (2, 2)	.2	7, 8	16
T(1, 2)	1.12	11-14	3	PE (3, 1)	.7	1, 2	17
T(2, 1)	7.46	1-4	4	PE (3, 2)	.3	7, 8	17
T(2, 2)	22.71	11-15	4	PE (4, 1)	.2	1, 2	18
T(3, 1)	5.	1, 2	5	PE (4, 2)	.8	7, 8	18
T(3, 2)	15.	11-13	5	PE (5, 1)	.1	1, 2	19
T(4, 1)	16.	1-3	6	PE (5, 2)	.9	7, 8	19
T(4, 2)	31.	11-13	6	FQ (1, 1)	3.	1, 2	20
T(5, 1)	14.	1-3	7	FQ (1, 2)	9.	11, 12	20
T(5, 2)	41.	11-13	7	FQ (2, 1)	4.	1, 2	21
T(6, 1)	10.	1-3	8	FQ (2, 2)	10.	11-13	21
T(6, 2)	30.	11-13	8	FQ (3, 1)	5.	1, 2	22
T(7, 1)	42.	1-3	9	FQ (3, 2)	11.	11-13	22
T(7, 2)	185.	11-13	9	FQ (4, 1)	4.	1, 2	23
T(8, 1)	15.	1-3	10	FQ (4, 2)	12.	11-13	23
T(8, 2)	25.	11-13	10	FQ (5, 1)	3.	1, 2	24
T(9, 1)	.0001	1-5	11	FQ (5, 2)	13.	11-13	24
T(9, 2)	76.8	11-14	11	FQ (6, 1)	1.	1, 2	25
T(10, 1)	.0001	1-5	12	FQ (6, 2)	2.	11, 12	25
T(10, 2)	1.	11, 12	12	FQ (7, 1)	4.	1, 2	26
PN (1)	.5	1, 2	13	FQ (7, 2)	6.	11, 12	26
PN (2)	.3	7, 8	13	FQ (8, 1)	5.	1, 2	27
PN (3)	.2	13, 14	13	FQ (8, 2)	8.	11, 12	27
PT (1)	.4	1, 2	14	FQ (9, 1)	7.	1, 2	28
PT (2)	.3	7, 8	14	FQ (9, 2)	11.	11-13	28
PT (3)	.15	13-15	14	FQ (10, 1)	9.	1, 2	29

TABLE 4. INFORMATION RETRIEVAL SYSTEM SIMULATION EXAMPLE - INPUT DATA FOR MOD II (Cont'd)

Variable	Value	Card Column	Card Number	Variable	Value	Card Column	Card Number
FQ (10,2)	14.	11-13	29	PQC(6, 1)	0		45
PQF(1,1)	5.	1, 2	30	PQC(6, 2)	0		45
PQF(1,2)	15.	11-13	30	PQC(7, 1)	0		46
PQF(2,1)	5.	1, 2	31	PQC(7, 2)	0		46
PQF(2,2)	15.	11-13	31	PQC(8, 1)	0		47
PQF(3,1)	5.	1, 2	32	PQC(8, 2)	0		47
PQF(3,2)	15.	11-13	32	PQC(9, 1)	0		48
PQF(4,1)	5.	1, 2	33	PQC(9, 2)	0		48
PQF(4,2)	15.	11-13	33	PQC(10, 1)	0		49
PQF(5,1)	5.	1, 2	34	PQC(10, 2)	0		49
PQF(5,2)	15.	11-13	34	CQC(1, 1)	5.	1, 2	50
PQF(6,1)	5.	1, 2	35	CQC(1, 2)	10.	11-13	50
PQF(6,2)	15.	11-13	35	CQC(2, 1)	5.	1, 2	51
PQF (7,1)	5.	1, 2	36	CQC(2, 2)	10.	11-13	51
PQF (7,2)	15.	11-13	36	CQC 3, 1)	5.	1, 2	52
PQF(8,1)	5.	1, 2	37	CQC(3, 2)	10.	11-13	52
PQF(8,2)	15.	11-13	37	CQC(4, 1)	5.	1, 2	53
PQF(9,1)	5.	1, 2	38	CQC(4, 2)	10.	11-13	53
PQF(9,2)	15.	11-13	38	CQC(5, 1)	5.	1, 2	54
PQF(10,1)	5.	1, 2	39	CQC (5, 2)	10.	11-13	54
PQF(10,2)	15.	11-13	39	CQC(6, 1)	0		55
PQC(1,1)	5.	1, 2	40	CQC(6, 2)	0		55
PQC(1,2)	20.	11-13	40	CQC(7, 1)	0		56
PQC (2,1)	5.	1, 2	41	CQC(7, 2)	0		56
PQC (2,2)	20.	11-13	41	CQC(8, 1)	0		57
PQC(3,1)	5.	1, 2	42	CQC(8, 2)	0		57
PQC(3,2)	20.	11-13	42	CQC(9, 1)	0		58
PQC(4,1)	5.	1, 2	43	CQC(9, 2)	0		58
PQC(4,2)	20.	11-13	43	CQC(10, 1)	0		59
PQC(5,1)	5.	1, 2	44	CQC(10, 2)	0		59
PQC(5,2)	20.	11-13	44	EQ(1, 1)	0		60

TABLE 4. INFORMATION RETRIEVAL SYSTEM SIMULATION EXAMPLE - INPUT DATA FOR MOD II (Cont'd)

Variable	Value	Card Column	Card Number	Variable	Value	Card Column	Card Number
EQ (1, 2)	0		60	ED (7, 1)	0		76
EQ (2, 1)	0		61	ED (7, 2)	0		76
EQ (2, 2)	0		61	ED (8, 1)	0		77
EQ (3, 1)	0		62	ED (8, 2)	0		77
EQ (3, 2)	0		62	ED (9, 1)	0		78
EQ (4, 1)	0		63	ED (9, 2)	0		78
EQ (4, 2)	0		63	ED (10, 1)	0		79
EQ (5, 1)	0		64	ED (10, 2)	0		79
EQ (5, 2)	0		64	EQC (1, 1)	3.	1, 2	80
EQ (6, 1)	5.	1, 2	65	EQC (1, 2)	10.	11-13	80
EQ (6, 2)	15.	11-13	65	EQC (2, 1)	4.	1, 2	81
EQ (7, 1)	5.	1, 2	66	EQC (2, 2)	10.	11-13	81
EQ (7, 2)	15.	11-13	66	EQC (3, 1)	5.	1, 2	82
EQ (8, 1)	5.	1, 2	67	EQC (3, 2)	10.	11-13	82
EQ (8, 2)	15.	11-13	67	EQC (4, 1)	6.	1, 2	83
EQ (9, 1)	5.	1, 2	68	EQC (4, 2)	10.	11-13	83
EQ (9, 2)	15.	11-13	68	EQC (5, 1)	7.	1, 2	84
EQ (10, 1)	5.	1, 2	69	EQC (5, 2)	10.	11-13	84
EQ (10, 2)	15.	11-13	69	EQC (6, 1)	0		85
ED (1, 1)	0		70	EQC (6, 2)	0		85
ED (1, 2)	0		70	EQC (7, 1)	0		86
ED (2, 1)	0		71	EQC (7, 2)	0		86
ED (2, 2)	0		71	EQC (8, 1)	0		87
ED (3, 1)	0		72	EQC (8, 2)	0		87
ED (3, 2)	0		72	EQC (9, 1)	0		88
ED (4, 1)	0		73	EQC (9, 2)	0		88
ED (4, 2)	0		73	EQC (10, 1)	0		89
ED (5, 1)	0		74	EQC (10, 2)	0		89
ED (5, 2)	0		74	TQT (1, 1)	0		90
ED (6, 1)	0		75	TQT (1, 2)	0		90
ED (6, 2)	0		75	TQT (2, 1)	0		91

TABLE 4. INFORMATION RETRIEVAL SYSTEM SIMULATION EXAMPLE - INPUT DATA FOR MOD II (Cont'd)

Variable	Value	Card Column	Card Number
TQT (2, 2)	0		91
TQT (3, 1)	0		92
TQT (3, 2)	0		92
TQT (4, 1)	0		93
TQT (4, 2)	0		93
TQT (5, 1)	0		94
TQT (5, 2)	0		94
TQT (6, 1)	0		95
TQT (6, 2)	0		95
TQT (7, 1)	0		96
TQT (7, 2)	0		96
TQT (8, 1)	0		97
TQT (8, 2)	0		97
TQT (9, 1)	0		98
TQT (9, 2)	0		98
TQT (10, 1)	0		99
TQT (10, 2)	0		99
DE (1)	.6	1, 2	100
TO (1, 1)	2.	7, 8	100
TO (1, 2)	12.	17-19	100
DE (2)	.1	1, 2	101
TO (2, 1)	10.	7-9	101
TO (2, 2)	35.	17-19	101
DE (3)	.3	1, 2	102
TO (3, 1)	5.	7, 8	102
TO (3, 2)	20.	17-19	102
A	.01	1-3	103
B	.008	7-10	103
C	.001	13-16	103

APPENDIX B

SIMULATION OUTPUT EXAMPLE

The following nine pages are a copy of the computer output from the Mod II simulation, using the input data given in Appendix A, Table 4. The headings for each data column are as follows:

A Iteration number

B Response time

C Number of queries in the iteration

D Query data

D1 Number of times query type 1 was used in the iteration

D2 Number of times query type 2 was used in the iteration

D3 Number of times query type 3 was used in the iteration

D4 Number of times query type 4 was used in the iteration

D5 Number of times query type 5 was used in the iteration

E Input device data

E1 Number of times input device 1 was used in the iteration

E2 Number of times input device 2 was used in the iteration

F Output device data

F1 Number of times output device 1 was used in the iteration

F2 Number of times output device 2 was used in the iteration

F3 Number of times output device 3 was used in the iteration

This output example will be discussed further in Chapter 3.

A	B	C	D					E		F		
			D1	D2	D3	D4	D5	E1	E2	F1	F2	F3
1	467.56	2	2	0	0	0	0	2	0	1	1	0
2	325.22	1	1	0	0	0	0	1	0	0	1	0
3	230.42	1	1	0	0	0	0	1	0	0	0	1
4	302.88	1	0	1	0	0	0	1	0	0	0	1
5	374.95	2	0	0	1	0	1	1	1	1	1	0
6	290.75	1	1	0	0	0	0	1	0	0	0	1
7	297.14	3	1	1	1	0	0	0	3	2	0	1
8	390.63	2	2	0	0	0	0	2	0	2	0	0
9	396.12	2	0	1	1	0	0	1	1	1	0	1
10	[illegible]	[illegible]	1	0	1	0	0	[illegible]	0	[illegible]	1	0
11	341.57	2	1	1	0	0	0	1	1	2	0	0
12	252.13	1	1	0	0	0	0	1	0	1	0	0
13	357.26	1	0	1	0	0	0	1	0	1	0	0
14	361.96	3	1	1	1	0	0	2	1	1	1	1
15	318.77	1	0	0	0	1	0	0	1	1	0	0
16	193.38	1	0	1	0	0	0	1	0	1	0	0
17	403.77	2	0	2	0	0	0	2	0	1	0	1
18	319.24	2	1	1	0	0	0	2	0	2	0	0
19	300.19	1	0	0	1	0	0	1	0	1	0	0
20	284.42	1	1	0	0	0	0	1	0	1	0	0
21	480.92	3	1	0	1	1	0	2	1	2	1	0
22	291.94	1	1	0	0	0	0	1	0	1	0	0
23	399.95	2	0	0	1	1	0	1	1	1	0	1
24	468.73	3	1	2	0	0	0	3	0	1	2	0
25	260.76	1	0	0	0	0	1	0	1	1	0	0
26	378.71	1	0	1	0	0	0	1	0	0	1	0
27	352.83	2	1	0	1	0	0	2	0	1	0	1
28	268.01	1	0	0	0	0	1	0	1	1	0	0
29	287.01	2	2	0	0	0	0	2	0	2	0	0
30	348.05	2	1	0	1	0	0	1	1	1	0	1
31	423.89	2	0	2	0	0	0	2	0	0	1	1
32	415.32	3	2	1	0	0	0	2	1	3	0	0
33	260.48	1	0	0	1	0	0	1	0	1	0	0
34	267.59	1	0	1	0	0	0	1	0	1	0	0
35	468.59	3	2	0	1	0	0	3	0	1	0	2
36	355.99	2	2	0	0	0	0	2	0	1	1	0
37	390.63	2	1	0	1	0	0	2	0	2	0	0
38	234.24	1	0	1	0	0	0	1	0	0	1	0
39	433.00	2	1	1	0	0	0	2	0	1	0	1
40	406.57	3	2	1	0	0	0	2	1	1	2	0
41	160.60	1	0	1	0	0	0	1	0	1	0	0
42	298.44	1	0	0	1	0	0	1	0	1	0	0
43	395.58	2	2	0	0	0	0	1	1	2	0	0
44	234.82	1	0	1	0	0	0	1	0	1	0	0
45	508.64	3	0	2	0	0	1	2	1	2	0	1
46	363.59	1	1	0	0	0	0	1	0	0	0	1
47	241.89	1	0	1	0	0	0	1	0	1	0	0
48	244.17	1	0	0	1	0	0	1	0	0	0	1
49	312.46	2	1	0	0	0	1	1	1	2	0	0
50	313.37	1	1	0	0	0	0	0	1	0	0	1
51	325.85	3	3	0	0	0	0	2	1	2	0	1
52	415.76	2	1	0	1	0	0	2	0	2	0	0
53	259.55	1	1	0	0	0	0	0	1	1	0	0
54	274.50	2	1	1	0	0	0	2	0	1	0	1
55	374.10	2	1	0	1	0	0	2	0	0	1	1
56	452.11	2	1	1	0	0	0	1	1	1	0	1
57	293.14	1	0	1	0	0	0	0	1	1	0	0
58	219.90	1	0	1	0	0	0	1	0	0	1	0

A	B	C	D					E		F		
			D1	D2	D3	D4	D5	E1	E2	F1	F2	F3
59	303.11	1	1	0	0	0	0	1	0	0	0	1
60	382.73	2	2	0	0	0	0	2	0	0	0	2
61	264.35	1	0	1	0	0	0	0	1	1	0	0
62	460.37	2	1	1	0	0	0	2	0	2	0	0
63	193.75	1	1	0	0	0	0	1	0	1	0	0
64	326.51	2	1	0	1	0	0	1	1	2	0	0
65	321.66	1	0	1	0	0	0	1	0	1	0	0
66	469.04	3	2	1	0	0	0	3	0	3	0	0
67	540.06	3	1	2	0	0	0	2	1	1	0	2
68	302.91	1	0	0	1	0	0	1	0	0	0	1
69	217.76	1	1	0	0	0	0	1	0	0	0	1
70	268.73	1	0	1	0	0	0	1	0	1	0	0
71	273.07	1	0	0	0	1	0	0	1	1	0	0
72	302.87	1	0	1	0	0	0	1	0	0	1	0
73	487.84	3	0	0	1	2	0	2	1	0	1	2
74	475.65	2	1	1	0	0	0	2	0	1	1	0
75	325.31	1	0	1	0	0	0	1	0	1	0	0
76	327.55	2	1	0	1	0	0	2	0	1	0	1
77	235.33	1	0	0	0	0	1	0	1	1	0	0
78	270.94	1	1	0	0	0	0	1	0	0	0	1
79	211.94	1	1	0	0	0	0	1	0	0	0	1
80	275.23	1	0	1	0	0	0	1	0	1	0	0
81	312.24	3	2	0	1	0	0	2	1	2	1	0
82	329.36	1	0	0	0	1	0	0	1	1	0	0
83	377.64	2	1	1	0	0	0	2	0	2	0	0
84	255.32	2	2	0	0	0	0	2	0	1	1	0
85	260.79	1	0	0	0	0	1	0	1	1	0	0
86	179.86	1	1	0	0	0	0	1	0	1	0	0
87	234.85	1	0	1	0	0	0	0	1	0	0	1
88	198.47	1	0	0	0	1	0	0	1	1	0	0
89	398.92	3	2	0	0	1	0	2	1	2	1	0
90	300.57	1	0	1	0	0	0	1	0	1	0	0
91	446.94	2	1	1	0	0	0	2	0	1	0	1
92	254.53	2	0	1	0	1	0	1	1	2	0	0
93	524.86	3	1	1	1	0	0	2	1	1	2	0
94	241.86	1	0	0	0	1	0	0	1	1	0	0
95	298.91	1	1	0	0	0	0	0	1	0	0	1
96	314.46	1	0	0	1	0	0	1	0	1	0	0
97	168.39	1	0	1	0	0	0	1	0	1	0	0
98	222.81	1	0	1	0	0	0	1	0	0	0	1
99	370.57	2	1	1	0	0	0	2	0	1	1	0
100	388.78	2	1	1	0	0	0	2	0	2	0	0
101	316.96	2	0	2	0	0	0	2	0	2	0	0
102	428.13	3	1	1	1	0	0	2	1	1	0	2
103	277.50	1	1	0	0	0	0	1	0	0	0	1
104	457.35	2	0	1	1	0	0	2	0	2	0	0
105	409.87	2	2	0	0	0	0	2	0	1	1	0
106	488.87	3	1	1	0	1	0	2	1	1	0	2
107	501.84	3	1	0	0	0	2	1	2	1	0	2
108	328.71	2	1	1	0	0	0	2	0	2	0	0
109	470.42	3	0	1	1	1	0	2	1	2	0	1
110	330.14	2	1	1	0	0	0	2	0	1	0	1
111	288.05	1	0	0	1	0	0	1	0	1	0	0
112	408.05	3	3	0	0	0	0	3	0	0	1	2
113	289.91	1	0	0	1	0	0	1	0	1	0	0
114	391.27	3	2	1	0	0	0	3	0	3	0	0
115	194.23	1	0	0	0	1	0	0	1	0	0	1
116	255.53	1	1	0	0	0	0	1	0	0	0	1

A	B	C	D1	D2	D3	D4	D5	E1	E2	F1	F2	F3
117	335.33	2	2	0	0	0	0	1	1	1	0	1
118	369.86	1	1	0	0	0	0	1	0	1	0	0
119	470.17	2	0	0	1	1	0	1	1	1	0	1
120	384.07	3	1	1	0	1	0	2	1	2	1	0
121	363.47	2	0	1	1	0	0	2	0	2	0	0
122	179.41	1	1	0	0	0	0	1	0	1	0	0
123	226.37	2	0	1	0	1	0	0	2	1	0	1
124	342.37	3	0	1	1	1	0	2	1	3	0	0
125	404.61	2	0	0	1	1	0	1	1	2	0	0
126	169.99	1	1	0	0	0	0	1	0	1	0	0
127	369.26	2	1	0	1	0	0	2	0	0	1	1
128	266.45	1	0	1	0	0	0	1	0	1	0	0
129	229.70	1	0	0	0	1	0	0	1	0	0	1
130	433.53	2	0	1	1	0	0	2	0	2	0	0
131	310.95	1	0	0	1	0	0	1	0	0	0	1
132	337.02	2	0	2	0	0	0	1	1	0	1	1
133	242.63	1	0	0	0	1	0	0	1	0	0	1
134	298.77	2	1	0	1	0	0	1	1	2	0	0
135	169.93	1	1	0	0	0	0	0	1	1	0	0
136	212.10	1	0	0	0	1	0	0	1	0	0	1
137	266.29	1	0	1	0	0	0	1	0	0	0	1
138	234.51	1	0	1	0	0	0	1	0	1	0	0
139	390.16	3	0	0	3	0	0	2	1	2	1	0
140	326.69	2	1	0	1	0	0	1	1	1	0	1
141	232.14	1	1	0	0	0	0	1	0	0	0	1
142	205.69	1	1	0	0	0	0	1	0	1	0	0
143	371.05	2	1	1	0	0	0	2	0	2	0	0
143	524.19	3	2	1	0	0	0	3	0	3	0	0
144	352.66	1	1	0	0	0	0	1	0	1	0	0
145	285.94	1	0	0	1	0	0	1	0	0	1	0
146	306.54	1	0	1	0	0	0	1	0	1	0	0
147	275.10	1	0	1	0	0	0	1	0	1	0	0
148	208.51	1	1	0	0	0	0	1	0	0	0	1
149	289.68	1	0	1	0	0	0	0	1	0	0	1
150	226.39	1	1	0	0	0	0	1	0	1	0	0
151	284.05	2	2	0	0	0	0	1	1	2	0	0
152	329.35	2	0	0	1	1	0	1	1	2	0	0
153	210.72	1	0	1	0	0	0	0	1	1	0	0
154	299.16	1	0	0	1	0	0	1	0	0	0	1
155	367.17	2	1	0	0	1	0	1	1	1	0	1
156	245.28	1	0	0	0	0	1	0	1	1	0	0
157	297.62	1	0	1	0	0	0	1	0	1	0	0
158	323.28	2	0	0	2	0	0	1	1	2	0	0
159	515.16	3	0	0	1	1	1	2	1	2	0	1
160	249.47	2	0	1	1	0	0	1	1	2	0	0
161	270.57	1	0	1	0	0	0	1	0	1	0	0
162	417.30	2	1	1	0	0	0	2	0	1	0	1
163	452.27	3	2	1	0	0	0	2	1	0	0	3
164	414.8[illegible]	3	1	0	1	0	1	2	1	3	0	0
165	360.94	2	1	1	0	0	0	1	1	1	0	1
166	452.50	2	0	0	1	1	0	0	2	0	1	1
167	251.38	1	0	1	0	0	0	1	0	0	0	1
168	310.81	1	0	0	1	0	0	0	1	0	0	1
169	283.24	1	1	0	0	0	0	1	0	1	0	0
170	413.69	2	2	0	0	0	0	2	0	1	1	0
171	265.17	1	1	0	0	0	0	1	0	0	1	0
172	294.19	2	1	1	0	0	0	2	0	1	1	0
173	517.26	3	0	1	1	1	0	2	1	3	0	0

A	B	C	D					E		F		
			D1	D2	D3	D4	D5	E1	E2	F1	F2	F3
174	324.34	1	0	0	0	0	1	0	1	0	0	1
175	307.73	2	1	0	0	0	1	0	2	1	0	1
176	279.06	1	0	0	0	1	0	0	1	1	0	0
177	374.93	3	1	1	1	0	0	2	1	3	0	0
178	390.13	2	1	1	0	0	0	2	0	1	0	1
179	235.15	1	1	0	0	0	0	1	0	1	0	0
180	258.30	1	0	1	0	0	0	1	0	0	0	1
181	233.12	1	0	1	0	0	0	1	0	1	0	0
182	304.84	1	1	0	0	0	0	0	1	1	0	0
183	300.82	1	0	0	0	0	1	0	1	0	1	0
184	285.91	1	0	1	0	0	0	1	0	0	0	1
185	376.59	2	1	1	0	0	0	1	1	1	0	1
186	239.64	1	0	0	1	0	0	1	0	1	0	0
187	234.78	2	0	0	0	2	0	0	2	1	0	1
188	347.63	2	0	0	1	1	0	0	2	1	0	1
189	286.97	1	0	1	0	0	0	1	0	1	0	0
190	451.76	3	0	1	1	1	0	1	2	2	0	1
191	278.37	1	0	1	0	0	0	1	0	1	0	0
192	304.35	2	1	1	0	0	0	2	0	2	0	0
193	240.67	1	0	0	0	1	0	0	1	0	0	1
194	258.11	2	0	2	0	0	0	2	0	0	0	2
195	432.79	3	1	1	0	1	0	2	1	1	0	2
196	282.55	1	0	1	0	0	0	0	1	1	0	0
197	428.38	3	2	1	0	0	0	2	1	1	0	2
198	497.59	3	1	1	0	1	0	2	1	2	1	0
199	347.48	1	1	0	0	0	0	1	0	0	1	0
200	190.09	1	1	0	0	0	0	1	0	0	0	1
201	295.59	1	0	1	0	0	0	1	0	1	0	0
202	377.94	3	1	1	0	0	1	2	1	2	0	1
203	284.18	1	0	1	0	0	0	1	0	0	0	1
204	341.28	3	0	2	1	0	0	2	1	2	0	1
205	258.36	1	0	0	0	1	0	0	1	1	0	0
206	158.80	1	1	0	0	0	0	0	1	0	0	1
207	251.18	1	0	1	0	0	0	0	1	1	0	0
208	345.05	1	1	0	0	0	0	1	0	0	0	1
209	327.14	1	0	1	0	0	0	1	0	1	0	0
210	220.80	1	0	1	0	0	0	1	0	0	1	0
211	287.04	1	0	1	0	0	0	1	0	1	0	0
212	175.60	1	0	1	0	0	0	0	1	1	0	0
213	292.04	1	1	0	0	0	0	1	0	1	0	0
214	249.80	1	0	0	1	0	0	1	0	0	0	1
215	438.91	2	0	2	0	0	0	1	1	1	1	0
216	465.15	2	2	0	0	0	0	2	0	1	1	0
217	233.57	1	1	0	0	0	0	1	0	0	0	1
218	269.59	1	0	0	1	0	0	1	0	0	1	0
219	383.38	3	2	1	0	0	0	1	2	1	0	2
220	428.29	3	2	1	0	0	0	2	1	3	0	0
221	229.16	1	0	0	1	0	0	1	0	0	1	0
222	327.82	1	0	1	0	0	0	1	0	0	0	1
223	325.35	1	1	0	0	0	0	1	0	0	1	0
224	189.24	1	0	0	0	0	1	0	1	1	0	0
225	460.42	3	1	1	1	0	0	3	0	3	0	0
226	453.67	3	2	0	0	1	0	3	0	3	0	0
227	220.69	1	0	1	0	0	0	0	1	1	0	0
228	263.51	1	1	0	0	0	0	1	0	0	0	1
229	291.04	1	1	0	0	0	0	1	0	1	0	0
230	442.05	2	2	0	0	0	0	2	0	0	1	1
231	412.90	2	0	1	1	0	0	2	0	0	0	2

A	B	C	D					E		F		
			D1	D2	D3	D4	D5	E1	E2	F1	F2	F3
232	396.72	2	2	0	0	0	0	2	0	1	0	1
233	295.43	1	1	0	0	0	0	1	0	1	0	0
234	171.14	1	0	1	0	0	0	1	0	1	0	0
235	344.78	2	1	1	0	0	0	2	0	1	0	1
236	413.49	3	2	1	0	0	0	2	1	2	0	1
237	396.51	2	0	1	1	0	0	1	1	2	0	0
238	429.52	2	1	1	0	0	0	1	1	1	0	1
239	345.65	1	0	1	0	0	0	0	1	1	0	0
240	322.06	2	1	1	0	0	0	2	0	1	0	1
241	410.58	2	0	1	1	0	0	2	0	2	0	0
242	298.01	1	0	1	0	0	0	1	0	0	0	1
243	170.91	1	0	0	0	1	0	0	1	1	0	0
244	310.53	1	0	1	0	0	0	0	1	1	0	0
245	485.55	3	2	0	0	0	1	3	0	1	1	1
246	250.79	1	0	1	0	0	0	1	0	0	0	1
247	241.22	1	0	1	0	0	0	0	1	0	0	1
248	245.18	1	1	0	0	0	0	1	0	1	0	0
249	264.32	1	0	0	0	0	1	1	0	1	0	0
250	451.01	3	1	1	0	1	0	2	1	3	0	0
251	398.00	2	1	1	0	0	0	1	1	2	0	0
252	332.70	1	1	0	0	0	0	1	0	1	0	0
253	333.11	2	1	0	0	0	1	1	1	1	1	0
254	541.78	3	3	0	0	0	0	2	1	1	1	1
255	497.95	3	1	2	0	0	0	3	0	1	0	2
256	381.77	3	1	1	0	0	1	2	1	1	0	2
257	339.51	1	1	0	0	0	0	1	0	1	0	0
258	260.04	1	1	0	0	0	0	1	0	0	0	1
259	398.74	2	2	0	0	0	0	2	0	0	1	1
260	277.04	2	1	1	0	0	0	2	0	0	0	2
261	266.52	2	1	0	1	0	0	0	2	1	0	1
262	314.11	2	0	2	0	0	0	2	0	2	0	0
263	426.21	3	1	1	0	1	0	2	1	3	0	0
264	331.11	1	0	1	0	0	0	0	1	0	0	1
265	515.31	3	1	1	1	0	0	3	0	3	0	0
266	469.01	3	2	1	0	0	0	3	0	1	1	1
267	164.04	1	1	0	0	0	0	1	0	1	0	0
268	449.30	2	2	0	0	0	0	1	1	1	1	0
269	354.97	3	2	0	[illegible]	0	1	2	1	2	0	1
270	507.55	3	0	2	0	0	1	2	1	1	0	2
271	333.78	1	0	0	1	0	0	1	0	1	0	0
272	351.49	2	0	2	0	0	0	2	0	1	1	0
273	380.62	3	0	1	0	1	1	1	2	2	0	1
274	301.09	2	0	2	0	0	0	2	0	2	0	0
275	342.82	1	1	0	0	0	0	1	0	1	0	0
276	429.38	3	1	2	0	0	0	3	0	3	0	0
277	396.68	3	3	0	0	0	0	2	1	3	0	0
278	358.11	3	2	0	0	1	0	2	1	3	0	0
279	243.68	1	0	0	0	1	0	1	0	1	0	0
280	350.16	3	0	1	2	0	0	1	2	3	0	0
281	298.78	1	1	0	0	0	0	1	0	0	0	1
282	201.99	1	0	1	0	0	0	0	1	1	0	0
283	291.09	1	0	1	0	0	0	1	0	1	0	0
284	303.78	1	0	1	0	0	0	1	0	1	0	0
285	198.22	1	0	1	0	0	0	1	0	1	0	0
286	311.12	2	1	1	0	0	0	1	1	0	1	1
287	355.16	2	1	1	0	0	0	2	0	1	0	1
288	239.84	1	1	0	0	0	0	1	0	1	0	0
289	215.73	1	0	1	0	0	0	1	0	1	0	0

A	B	C	D					E		F		
			D1	D2	D3	D4	D5	E1	E2	F1	F2	F3
290	380.36	3	1	2	0	0	0	3	0	1	0	2
291	298.85	1	0	1	0	0	0	0	1	1	0	0
292	297.61	1	1	0	0	0	0	1	0	1	0	0
293	385.03	2	0	0	2	0	0	1	1	1	1	0
294	213.44	1	0	0	1	0	0	0	1	0	0	1
295	443.44	2	0	1	1	0	0	2	0	2	0	0
296	441.85	3	0	1	0	1	1	1	2	1	1	1
297	312.45	1	0	0	0	0	1	0	1	0	1	0
298	564.32	3	1	1	1	0	0	2	1	2	1	0
299	240.97	1	1	0	0	0	0	1	0	1	0	0
300	199.04	1	0	0	0	1	0	1	0	1	0	0
301	203.30	1	1	0	0	0	0	1	0	1	0	0
302	460.07	2	1	1	0	0	0	1	1	1	1	0
303	257.01	1	0	1	0	0	0	0	1	0	1	0
304	441.60	3	1	1	1	0	0	2	1	3	0	0
305	418.98	2	0	1	0	0	1	1	1	2	0	0
306	349.39	1	1	0	0	0	0	1	0	1	0	0
307	423.98	3	1	0	2	0	0	3	0	1	0	2
308	148.76	1	0	0	0	0	1	0	1	1	0	0
309	225.29	1	0	1	0	0	0	1	0	1	0	0
310	339.09	2	0	0	0	2	0	0	2	1	1	0
311	352.79	2	1	0	1	0	0	2	0	1	1	0
312	344.83	3	2	0	1	0	0	2	1	3	0	0
313	384.84	2	2	0	0	0	0	2	0	1	0	1
314	298.47	1	1	0	0	0	0	1	0	1	0	0
315	502.91	3	0	2	1	0	0	3	0	1	2	0
316	442.44	3	1	0	1	1	0	0	3	1	0	2
317	350.36	1	1	0	0	0	0	1	0	1	0	0
318	424.58	3	1	1	1	0	0	2	1	2	1	0
319	395.22	2	0	0	2	0	0	1	1	1	0	1
320	350.93	1	0	1	0	0	0	1	0	1	0	0
321	194.74	1	1	0	0	0	0	1	0	1	0	0
322	391.49	2	2	0	0	0	0	2	0	1	0	1
323	462.77	3	1	2	0	0	0	1	2	2	0	1
324	431.39	3	1	1	0	1	0	3	0	1	0	2
325	261.67	1	0	0	0	1	0	0	1	0	1	0
326	337.09	2	1	1	0	0	0	2	0	2	0	0
327	379.64	3	2	1	0	0	0	3	0	2	0	1
328	322.50	2	1	1	0	0	0	2	0	1	1	0
329	330.36	1	1	0	0	0	0	1	0	0	0	1
330	270.63	2	1	0	0	1	0	1	1	1	0	1
331	296.71	1	0	0	0	1	0	0	1	0	1	0
332	288.15	1	0	1	0	0	0	1	0	1	0	0
333	417.65	3	0	1	0	1	1	1	2	2	0	1
334	341.53	2	1	0	0	1	0	1	1	2	0	0
335	330.66	1	1	0	0	0	0	1	0	0	0	1
336	358.22	2	1	1	0	0	0	2	0	1	0	1
337	288.53	1	1	0	0	0	0	0	1	1	0	0
338	454.00	3	1	2	0	0	0	2	1	2	1	0
339	503.46	3	0	3	0	0	0	3	0	2	1	0
340	390.72	1	0	1	0	0	0	0	1	0	0	1
341	358.91	1	0	1	0	0	0	1	0	1	0	0
342	349.61	2	0	2	0	0	0	2	0	2	0	0
343	299.69	2	1	0	0	0	1	0	2	1	0	1
344	200.87	1	1	0	0	0	0	1	0	1	0	0
345	187.48	1	1	0	0	0	0	1	0	1	0	0
346	381.41	2	2	0	0	0	0	2	0	1	0	1
347	540.30	3	1	0	0	0	2	1	2	0	0	3

A	B	C	D					E		F		
			D1	D2	D3	D4	D5	E1	E2	F1	F2	F3
348	377.63	2	0	0	2	0	0	2	0	1	1	0
349	354.60	1	1	0	0	0	0	1	0	0	0	1
350	290.35	1	0	1	0	0	0	1	0	1	0	0
351	461.53	3	1	1	1	0	0	3	0	3	0	0
352	279.38	1	0	0	1	0	0	1	0	0	0	1
353	241.57	1	0	1	0	0	0	0	1	1	0	0
354	363.96	2	1	0	0	0	1	1	1	2	0	0
355	318.70	2	1	0	0	0	1	1	1	0	1	1
356	477.05	3	0	1	1	1	0	2	1	1	1	1
357	261.36	1	1	0	0	0	0	0	1	0	0	1
358	197.41	1	0	0	0	1	0	0	1	1	0	0
359	259.71	1	1	0	0	0	0	1	0	0	0	1
360	355.64	1	0	1	0	0	0	1	0	0	0	1
361	304.58	2	2	0	0	0	0	2	0	1	0	1
362	226.93	1	0	1	0	0	0	1	0	1	0	0
363	416.66	2	0	1	1	0	0	1	1	0	1	1
364	376.50	2	1	1	0	0	0	2	0	1	0	1
365	399.86	2	1	0	1	0	0	0	2	0	1	1
366	338.96	2	1	1	0	0	0	2	0	1	1	0
367	274.43	1	0	0	0	1	0	0	1	0	0	1
368	265.99	1	0	0	0	1	0	0	1	0	1	0
369	368.35	2	1	0	1	0	0	2	0	1	1	0
370	312.36	2	0	1	1	0	0	1	1	2	0	0
371	337.63	1	0	1	0	0	0	1	0	1	0	0
372	224.72	1	0	0	0	1	0	0	1	0	1	0
373	333.60	2	1	0	1	0	0	2	0	1	0	1
374	385.06	2	0	1	0	1	0	0	2	0	1	1
375	281.94	2	1	1	0	0	0	2	0	1	0	1
376	250.62	1	1	0	0	0	0	1	0	1	0	0
377	218.07	1	0	0	1	0	0	0	1	0	0	1
378	302.80	2	0	2	0	0	0	1	1	2	0	0
379	419.17	3	0	3	0	0	0	2	1	0	1	2
380	326.92	1	1	0	0	0	0	1	0	1	0	0
381	287.66	1	0	0	1	0	0	1	0	0	0	1
382	288.97	1	0	0	1	0	0	1	0	0	0	1
383	337.09	2	1	0	1	0	0	2	0	1	0	1
384	235.34	1	1	0	0	0	0	1	0	1	0	0
385	253.43	1	0	0	0	1	0	1	0	0	1	0
386	247.63	1	1	0	0	0	0	1	0	1	0	0
387	357.24	2	1	0	0	1	0	1	1	1	1	0
388	277.95	1	0	1	0	0	0	0	1	0	1	0
389	342.78	2	0	0	0	2	0	1	1	1	0	1
390	285.36	1	1	0	0	0	0	0	1	1	0	0
391	405.34	2	1	1	0	0	0	2	0	0	1	1
392	308.97	1	0	1	0	0	0	1	0	0	0	1
393	262.20	1	0	1	0	0	0	1	0	0	0	1
394	431.24	3	1	2	0	0	0	3	0	2	0	1
395	286.55	1	0	0	0	1	0	1	0	1	0	0
396	327.93	2	1	1	0	0	0	1	1	1	0	1
397	386.55	3	0	1	1	1	0	1	2	1	0	2
398	313.06	1	1	0	0	0	0	1	0	1	0	0
399	230.93	1	0	0	0	1	0	1	0	1	0	0
400	305.87	1	1	0	0	0	0	1	0	1	0	0
401	261.91	1	0	0	0	0	1	0	1	1	0	0
402	325.65	1	0	0	0	0	1	0	1	1	0	0
403	289.00	1	0	1	0	0	0	1	0	1	0	0
404	307.91	1	0	0	1	0	0	0	1	0	0	1
405	189.53	1	0	0	1	0	0	0	1	0	0	1

A	B	C	D					E		F		
			D1	D2	D3	D4	D5	E1	E2	F1	F2	F3
406	338.25	2	1	1	0	0	0	2	0	1	0	1
407	359.99	2	0	1	0	1	0	1	1	2	0	0
408	438.66	2	1	0	0	1	0	1	1	1	0	1
409	395.84	2	0	1	0	0	1	1	1	0	1	1
410	368.31	2	0	1	0	0	1	1	1	0	1	1
411	251.75	1	1	0	0	0	0	1	0	0	1	0
412	226.28	1	1	0	0	0	0	1	0	0	0	1
413	205.06	1	0	1	0	0	0	1	0	0	0	1
414	332.18	1	1	0	0	0	0	1	0	1	0	0
415	432.76	3	2	1	0	0	0	3	0	3	0	0
416	321.35	1	1	0	0	0	0	1	0	1	0	0
417	209.88	1	1	0	0	0	0	1	0	1	0	0
418	295.17	1	1	0	0	0	0	1	0	0	1	0
419	254.49	1	0	1	0	0	0	1	0	1	0	0
420	515.03	3	1	0	0	2	0	1	2	0	1	2
421	395.29	3	0	2	0	0	1	1	2	2	0	1
422	459.65	3	0	1	1	1	0	1	2	2	0	1
423	332.80	1	0	0	1	0	0	1	0	1	0	0
424	257.11	1	1	0	0	0	0	1	0	1	0	0
425	252.96	1	0	1	0	0	0	1	0	1	0	0
426	356.23	2	0	0	1	1	0	1	1	2	0	0
427	251.39	1	1	0	0	0	0	1	0	1	0	0
428	508.99	3	1	1	1	0	0	2	1	3	0	0
429	566.44	3	1	0	2	0	0	1	2	1	0	2
430	295.25	1	0	1	0	0	0	0	1	1	0	0
431	402.20	1	0	1	0	0	0	1	0	0	1	0
432	335.17	2	1	1	0	0	0	2	0	2	0	0
433	375.57	1	1	0	0	0	0	1	0	0	0	1
434	235.12	1	0	0	0	0	1	0	1	0	1	0
435	428.53	2	0	0	1	1	0	1	1	2	0	0
436	424.63	3	0	3	0	0	0	2	1	0	0	3
437	381.03	2	0	0	1	0	1	2	0	1	0	1
438	320.82	1	0	1	0	0	0	1	0	1	0	0
439	434.85	2	0	1	0	0	1	1	1	2	0	0
440	272.26	1	0	1	0	0	0	1	0	1	0	0
441	183.46	1	0	1	0	0	0	1	0	0	0	1
442	309.41	2	1	0	1	0	0	2	0	1	0	1
443	238.87	1	1	0	0	0	0	1	0	1	0	0
444	325.80	2	2	0	0	0	0	2	0	0	1	1
445	305.08	1	1	0	0	0	0	1	0	0	0	1
446	357.32	2	1	1	0	0	0	2	0	0	1	1
447	376.49	2	0	1	1	0	0	1	1	2	0	0
448	411.70	2	1	1	0	0	0	2	0	2	0	0
449	181.69	1	1	0	0	0	0	1	0	1	0	0
450	314.78	2	0	1	1	0	0	2	0	2	0	0
451	282.77	2	2	0	0	0	0	1	1	0	0	2
452	284.45	1	1	0	0	0	0	1	0	1	0	0
453	404.03	2	0	0	1	1	0	2	0	0	1	1
454	245.98	1	0	1	0	0	0	0	1	1	0	0
455	352.88	2	0	0	1	1	0	2	0	2	0	0
456	385.36	2	1	0	1	0	0	2	0	2	0	0
457	390.67	2	1	0	0	1	0	1	1	1	1	0
458	515.42	3	1	2	0	0	0	3	0	3	0	0
459	205.67	1	0	1	0	0	0	1	0	0	0	1
460	388.02	2	2	0	0	0	0	2	0	0	1	1
461	395.34	2	1	0	1	0	0	2	0	2	0	0
462	235.35	1	0	1	0	0	0	1	0	0	0	1
463	320.81	1	1	0	0	0	0	1	0	1	0	0

A	B	C	D					E		F		
			D1	D2	D3	D4	D5	E1	E2	F1	F2	F3
464	521.40	3	1	2	0	0	0	3	0	2	1	0
465	203.92	1	0	1	0	0	0	1	0	1	0	0
466	223.24	1	1	0	0	0	0	0	1	0	0	1
467	382.19	2	0	2	0	0	0	1	1	1	0	1
468	222.79	1	0	1	0	0	0	1	0	1	0	0
469	486.45	3	1	1	0	0	1	2	1	2	0	1
470	531.65	3	2	0	1	0	0	2	1	3	0	0
471	384.51	3	1	1	0	0	1	2	1	2	1	0
472	466.73	3	0	2	1	0	0	1	2	1	1	1
473	226.92	1	1	0	0	0	0	1	0	0	0	1
474	387.79	2	1	0	1	0	0	1	1	1	0	1
475	250.43	1	0	1	0	0	0	1	0	1	0	0
476	248.00	1	1	0	0	0	0	1	0	1	0	0
477	290.91	1	0	0	0	1	0	0	1	1	0	0
478	415.33	3	1	1	1	0	0	2	1	1	0	2
479	583.95	3	2	1	0	0	0	3	0	1	1	1
480	363.00	2	0	1	0	1	0	1	1	2	0	0
481	391.71	3	3	0	0	0	0	3	0	2	0	1
482	291.29	2	1	0	0	1	0	1	1	0	0	2
483	357.66	1	0	1	0	0	0	1	0	0	0	1
484	236.05	1	0	0	1	0	0	0	1	1	0	0
485	450.42	3	2	1	0	0	0	3	0	3	0	0
486	260.88	1	1	0	0	0	0	1	0	0	0	1
487	229.37	1	1	0	0	0	0	1	0	0	0	1
488	389.88	2	1	1	0	0	0	1	1	2	0	0
489	191.43	1	1	0	0	0	0	1	0	0	1	0
490	497.88	3	3	0	0	0	0	2	1	2	1	0
491	334.32	1	0	1	0	0	0	1	0	1	0	0
492	262.42	1	0	1	0	0	0	1	0	0	0	1
493	301.06	1	0	0	0	1	0	0	1	1	0	0
494	300.62	1	1	0	0	0	0	1	0	1	0	0
495	334.28	2	2	0	0	0	0	2	0	2	0	0
496	436.33	2	0	2	0	0	0	2	0	0	1	1
497	261.34	1	0	1	0	0	0	0	1	1	0	0
498	292.81	1	1	0	0	0	0	1	0	1	0	0
499	335.02	2	1	1	0	0	0	1	1	0	1	1
500	270.96	2	2	0	0	0	0	2	0	1	0	1

HRB

DISTRIBUTION LIST

Address	Number of copies
Technical Library Director Defense Res. & Eng. Room 3C-128, The Pentagon Washington, D.C. 20301	1
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20
Chief of Naval Research Department of the Navy Washington 25, D.C. Attn: Code 437, Information Systems Branch	2
Chief of Naval Operations OP-07T-12 Navy Department Washington 25, D.C.	1
Director, Naval Research Laboratory Technical Information Officer/Code 2000 Washington 25, D.C.	6
Commanding Officer, Office of Naval Research Navy #100, Fleet Post Office Box 39 New York, New York, 09599	10
Commanding Officer ONR Branch Office 207 West 24th Street New York 11, New York	1
Office of Naval Research Branch Office 495 Summer Street Boston, Massachusetts 02110	1
Naval Ordnance Laboratory White Oaks Silver Spring 19, Maryland Attn: Technical Library	1
David Taylor Model Basin Code 042 Technical Library Washington, D.C. 20007	1
Naval Electronics Laboratory San Diego 52, California Attn: Technical Library	1

DISTRIBUTION LIST (cont'd)

Address	Number of copies
Dr. Daniel Alpert, Director Coordinated Science Laboratory University of Illinois Urbana, Illinois	1
Air Force Cambridge Research Labs Laurence C. Hanscom Field Bedford, Massachusetts Attn: Research Library, CRMXL-R	1
U.S. Naval Weapons Laboratory Dahlgren, Virginia 22448 Attn: G.H. Gleissner/Code K-4/ Asst. Director for Computation	1
National Bureau of Standards Data Processing Systems Division Room 239, Bldg. 10 Attn: A.K. Smilow Washington 25, D.C.	1
George C. Francis Computing Lab., BRL Aberdeen Proving Ground, Maryland	1
Office of Naval Research Branch Office Chicago 230 N. Michigan Avenue Chicago, Illinois 60601	1
Commanding Officer ONR Branch Office 1030 E. Green Street Pasadena, California	1
Commanding Officer ONR Branch Office 1000 Geary Street SanfFrancisco 9, California	1
National Physical Laboratory Teddington, Middlesex England Attn: Dr. A.M. Uttley, Superintendent, Autonomics Division	1

DISTRIBUTION LIST (cont'd)

Address	Number of copies
Commanding Officer Harry Diamond Laboratories Attn: Library Washington, D. C. 20438	1
Commanding Officer and Director U.S. Naval Training Device Center Port Washington Long Island, New York Attn: Technical Library	1
Department of the Army Office of the Chief of Research & Development Pentagon, Room 3D442 Washington 25, D. C. Attn: Mr. L. H. Geiger	1
National Security Agency Fort George G. Meade, Maryland Attn: Librarian, C-332	1
Lincoln Laboratory Massachusetts Institute of Technology Lexington 73, Massachusetts Attn: Library	1
Mr. A. Bodian Navy Department OPNAV - OP-922 The Pentagon Washington 25, D. C.	1
HRB-Singer, Inc. Science Park State College, Pennsylvania Attn: Mr. Ralph Stevenson	1
Assistant Chief of Naval Operations Office of Naval Intelligence Washington 25, D. C. Attn: Assistant for Automation Coordination Code 92B4	1

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) 5 HRB-Singer, Inc.	2a. REPORT SECURITY CLASSIFICATION Unclassified
	2b. GROUP

3. REPORT TITLE
6 The Simulation and Evaluation of Information Retrieval Systems

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (Last name, first name, initial) 10 by
Conger, C. Richard

6. REPORT DATE 11 Apr[redacted]65	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS none
8a. CONTRACT [redacted] 15 Nonr 3818(00)	9a. ORIGINATOR'S REPORT NUMBER(S) 14 Rept. no. 352-R-17	
b. PROJECT NO.		
c	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
d.		

10. AVAILABILITY/LIMITATION NOTICES
Qualified requesters may obtain copies of this report from DDC.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research Main Navy Washington, D. C.

13. ABSTRACT

The assumptions, logic, and operations of an Information Retrieval System simulation model are described in this report. The model has been programmed in FORTRAN II for the IBM 1620. Two model variations, input and output examples, and some possibilities for improvement and development are given. Data reduction and analysis programs, operating upon the output of the simulation model, are explained and a set of sample outputs for these programs are given. Also included are listings of the latest programs developed.

DD FORM 1 JAN 64 1473

Unclassified
Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Computer Simulation, Computer Systems Components, Evaluation, Information Retrieval Systems, Mathematics, Models, Monte Carlo, Programming, Sampling, Simulation, Statistical Distributions, Statistics, System Analysis.						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ____________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ____________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ____________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms [illegible]

Best Available Copy